# CMS Draft Analysis Note

*The content of this note is intended for CMS internal use and distribution only*

2015/03/06
Head Id: 279796
Archive Id: 279655:279796M
Archive Date: 2015/03/06
Archive Tag: trunk

# Update on the search in the $lljj$ final states at $\sqrt{s}$ =8 TeV

J. Butler [1], H. Cheung [1], L. Uplegger [1], J. Nachtman [2], Y. Onel [2], H. Ogul [2], and P. Tan [2]

[1] Fermilab
[2] University of Iowa



## Abstract

We report an update on a search in the $lljj$ final states using the full 2012 CMS data at $\sqrt{s}$=8 TeV, which corresponds to an integrated luminosity of 19.7 fb$^{-1}$. Motivated by an potential "excess" spotted in 2011 CMS data we have systematically examined the $ll + jj$ mass spectrum using 2012 CMS data. This involves analyzing the mass spectrum in 9 statistically independent primary data regions defined by eletron and muon flavor combination and mass range of dileptons (above Z-pole, on Z-pole, and below Z-pole). In each selected data sample, we scan the $m_{lljj}$ mass spectrum from [700, 1440] GeV with a 10% ([0.9m, 1.1m]) sliding mass window, using MC simulations to estimate known SM backgrounds. In $ee$ and $e\mu$ channels, no significant local "excess" above SM expectation has been observed. In $\mu\mu$ channel above Z-pole region, we observe an excess around 1 TeV mass with a local significance of 3.7 $\sigma$ after including systematic uncertainties. The significance of this local "excess" is 4.1 $\sigma$ without systeamtic uncertainties. In the $\mu\mu$ below Z-pole region, small "excess" is also presented in the similar region as shown in above Z-pole sample. Additional studies and re-analyzing of 2011 data are still ongoing.

# 1 The Introduction

In this note, we present an update of a search in the $lljj$ final states using the full 2012 CMS data. At the beginning of 2011 LHC run, while studying other physics subjects, by accident we have noticed a moderate "excess" in the $\mu\mu jj$ final state around the mass range of 1 TeV using the first 1.1 fb$^{-1}$ CMS data, as shown in Fig. 1. The dominant standard model (SM) processes shown in Fig. 1 are Drell-Yan (DY) + jets and $t\bar{t}$ production. The follow-ups on this "excess" with additional 2011 CMS data have been documented as internal analysis notes [1, 2]. Comparing to the selections used to make Fig. 1, in notes [1, 2] we only slightly increased the leading jet $p_T$ threshold and all other selections remained to be same. The main motivation was to observe how the "excess" evolves with incoming data.

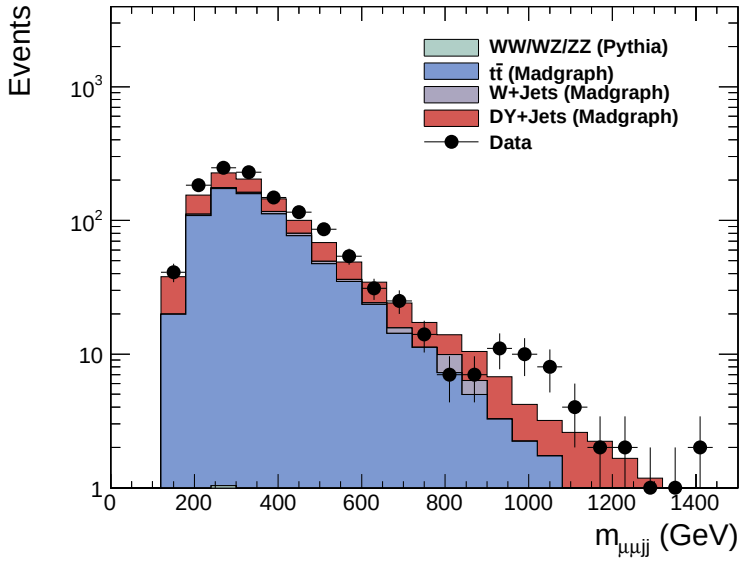


Figure 1: The $\mu\mu jj$ mass distribution obtained using the first 1.1 fb$^{-1}$ of the CMS 2011 data. All MC simulations are normalized to data luminosity using cross sections. The selection used to make this plot was intended for other physics studies.

Several groups have performed independent cross checks on the mass spectrum we have obtained and these are documented in this internal note [3]. The $\mu\mu jj$ invariance mass spectrum obtained by different groups are shown in Fig. 2, where the 2011 CMS data have been divided into Run2011A and Run2011B subsets with 2.3 fb$^{-1}$ and 2.7 fb$^{-1}$ integrated luminosity, respectively. In both data subsets, the $\mu\mu jj$ spectrum has been reproduced to very good extent following identical/similar selections. The relatively larger discrepancies between Darin's mass spectrum and the others are mainly due to the trigger matching, which is not applied in Darin's cross-check.

The "excess" is more obvious in 2011A than in 2011B. Preliminary studies [1, 2] show no evidence of "excess" in the $eejj$ and $e\mu jj$ final states. Motivated by this potential "excess", we have investigated this particular mass region ($\sim$ 1 TeV) in the $lljj$ final states using the full 2012 CMS data. This analysis note is served as a documentation of our latest findings. Our focus is on the off Z-pole same flavor dilepton channels ($\mu\mu jj$ and $eejj$). The general strategy is to rely on MC simulations for modeling the known SM processes and systematically examine the $lljj$ invariant mass spectrum. We use $e\mu$ data and on Z-pole same favor dilepton channels as control samples to validate of DY+Jets and $t\bar{t}$ simulations, which are the two dominant SM processes.

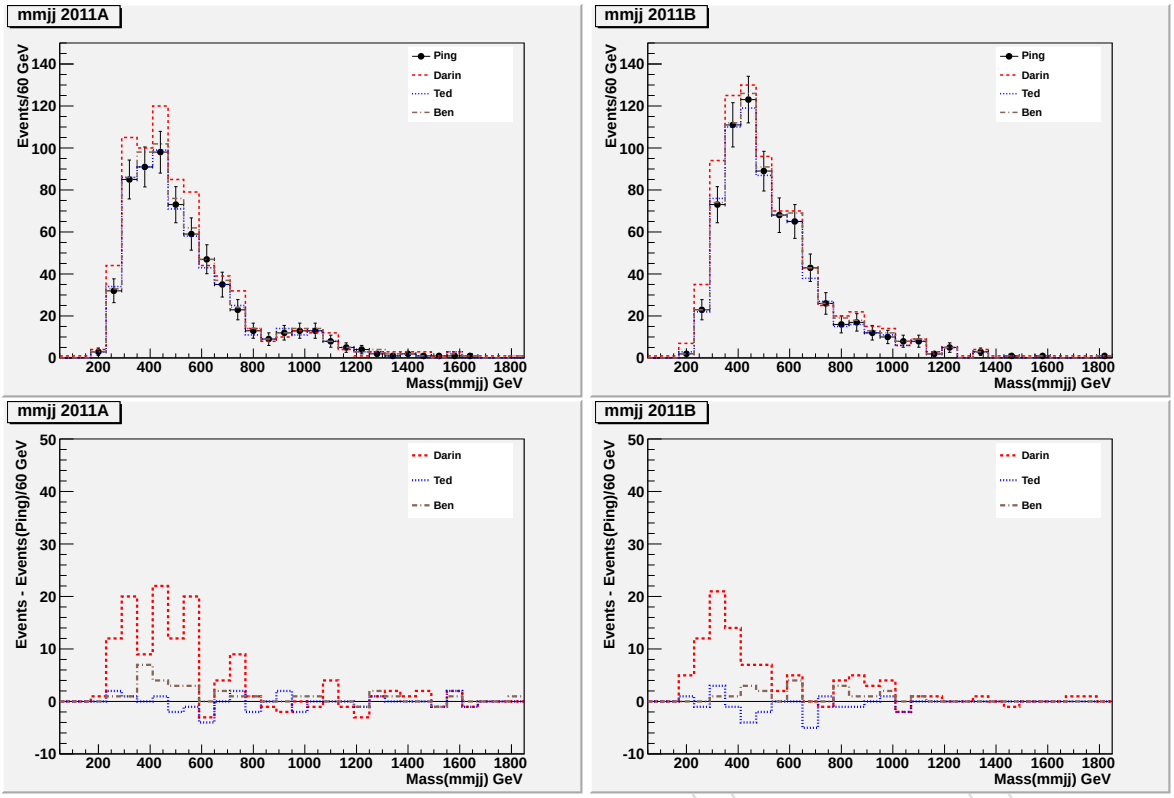


Figure 2: Comparisons of the $\mu\mu jj$ mass distributions among four independent analyses. Top) the mass spectrum from Run 2011A is shown in the left subfigure and the one for Run 2011B is shown on the right. Bottom) Differences among separate analyses are shown on the left subfigure and right subfigure for Ru 2011A dataset and Run2011B dataset, respectively.

One thing worthy to point out is that many searches involved $lljj$ final states at CMS are optimized for a particular new physics (NP) model, and somehow these analyses (E.g. Ref [4, 5]) have less sensitivities to the "excess" we are exploring here.

This note is organized as follows. In Section 2, we listed the data samples and Monte Carlo (MC) simulations used to model dominant SM processes. The certified objects used in this analysis are described in Section 3. Several corrections are applied to MC simulations to improve the data/MC agreement and they are described in details in Section 4 and some validation plots are also shown. The analysis is described in Section 5. The systematic uncertainties considered are discussed in Section 6 and the results are shown in Section 7. The summary of this analysis and the final conclusion are in Section 8. Cross checks of the results are listed in Appendix.

# 2 Data Sets

The data used in this analysis are shown in Table 1. We use the double lepton datasets as primary datasets for this analysis, and the single lepton datasets are used to determine lepton efficiencies and cross checks.

The good run/luminosity sections are selected based on,

Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON.txt

The following event filters are applied,

- "noscraping"
- "primaryVertexFilter"
- "HBHENoiseFilter"

and only events passing these filters are considered in the analysis. The global tag used to process data is "FT_53_V21_AN6::All". The luminosity is calculated using the "pixelLumiCalc.py" in the "V04-02-10" RecoLuminosity package.

Table 1: Summary of data samples and corresponding luminosities. The double lepton primary datasets are used as the default for this analysis.

| Dataset | Run Range | Trigger Path | Lumi. ($pb^{-1}$) |
|---|---|---|---|
| DoubleElectron | | | |
| /DoubleElectron/Run2012A-22Jan2013-v1/AOD | 190456-193621 | HLT_Ele17_*_Ele8* | 871.738 |
| /DoubleElectron/Run2012B-22Jan2013-v1/AOD | 193833-196531 | HLT_Ele17_*_Ele8* | 4401 |
| /DoubleElectron/Run2012C-22Jan2013-v1/AOD | 198022-203742 | HLT_Ele17_*_Ele8* | 7045 |
| /DoubleElectron/Run2012D-22Jan2013-v1/AOD | 203777-208686 | HLT_Ele17_*_Ele8* | 7366 |
| Total | | | 19683.738 |
| MuEG | | | |
| /MuEG/Run2012A-22Jan2013-v1/AOD | 190456-193621 | HLT_Mu17_Ele8* | 876.225 |
| /MuEG/Run2012B-22Jan2013-v1/AOD | 193833-196531 | HLT_Mu17_Ele8* | 4408 |
| /MuEG/Run2012C-22Jan2013-v1/AOD | 198022-203742 | HLT_Mu17_Ele8* | 7041 |
| /MuEG/Run2012D-22Jan2013-v1/AOD | 203777-208686 | HLT_Mu17_Ele8* | 7344 |
| Total | | | 19669.225 |
| DoubleMuon | | | |
| /DoubleMu/Run2012A-22Jan2013-v1/AOD | 190456-193621 | HLT_Mu17_Mu8* | 876.225 |
| /DoubleMuParked/Run2012B-22Jan2013-v1/AOD | 193833-196531 | HLT_Mu17_Mu8* | 4410 |
| /DoubleMuParked/Run2012C-22Jan2013-v1/AOD | 198022-203742 | HLT_Mu17_Mu8* | 7013 |
| /DoubleMuParked/Run2012D-22Jan2013-v1/AOD | 203777-208686 | HLT_Mu17_Mu8* | 7360 |
| Total | | | 19659.225 |
| SingleElectron | | | |
| /SingleElectron/Run2012A-22Jan2013-v1/AOD | 190456-193621 | HLT_Ele27_WP80* | 871.538 |
| /SingleElectron/Run2012B-22Jan2013-v1/AOD | 193833-196531 | HLT_Ele27_WP80* | 4395 |
| /SingleElectron/Run2012C-22Jan2013-v1/AOD | 198022-203742 | HLT_Ele27_WP80* | 7014 |
| /SingleElectron/Run2012D-22Jan2013-v1/AOD | 203777-208686 | HLT_Ele27_WP80* | 7357 |
| Total | | | 19637.538 |
| SingleMuon | | | |
| /SingleMu/Run2012A-22Jan2013-v1/AOD | 190456-193621 | HLT_IsoMu24_eta2p1* | 876.225 |
| /SingleMu/Run2012B-22Jan2013-v1/AOD | 193833-196531 | HLT_IsoMu24* | 4408 |
| /SingleMu/Run2012C-22Jan2013-v1/AOD | 198022-203742 | HLT_IsoMu24* | 7041 |
| /SingleMu/Run2012D-22Jan2013-v1/AOD | 203777-208686 | HLT_IsoMu24* | 7344 |
| Total | | | 19675 |

Table 2 shows the primary MC simulations used in this analysis. To simulate $t\bar{t}$ and single top *tW*-production, the POWHEG [6] event generator is used along with CT10 PDF model [7]. The DY+Jets and the di-boson (WW, WZ, ZZ) process are simulated using the MADGRAPH matrix-element event generator [8] and the CTEQ6L1 PDF model [9] has been used.All MC simulations are passed to PYTHIA (v.6.422) event generator [10] for shower simulation and modeling of

underlying event (UE). The TuneZ2* UE model has been used. The $\tau$ decay is simulated using the TAUOLA package [11] to take into account the $\tau$ polarization. The global tag used to process MC simulations is "START53_V27::All". Additional PYTHIA-based di-boson simulations and MADGRAPH-based $t\bar{t}$ simulations are used as cross checks. We use normalization cross sections listed in this twiki page,

https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSectionsat8TeV

and the PDG-averaged branching fractions are used for leptonic enriched samples.

Table 2: MC simulation and normalization cross sections used in this analysis. The normalization cross sections have taken into account the branching fractions for leptonic enriched samples using values from PDG [12]. Note that the DY+0Jets MC simulation is a subset of the inclusive DY+Jets sample.

.

| Type | Num. of Events | $\sigma$ (pb) | Luminosity ($fb^{-1}$) |
|---|---|---|---|
| $t\bar{t}$ (POWHEG) | 10783473 | 252.89 | 400.1 |
| DY+Jets (Madgraph) | 30435326 | 3531.89 | 8.6 |
| DY+0Jets (Madgraph) | 22016629 | 2554.81 | 8.6 |
| DY+1Jets (Madgraph) | 24045135 | 666.15 | 36.1 |
| DY+2Jets (Madgraph) | 21756105 | 216.70 | 100.4 |
| DY+3Jets (Madgraph) | 11003418 | 66.45 | 165.6 |
| DY+4Jets (Madgraph) | 6402800 | 27.78 | 230.5 |
| WW(2l2$\nu$) (Madgraph) | 1933225 | 5.94 | 325.7 |
| WZ(3l$\nu$) (Madgraph) | 2017967 | 1.11 | 1825.5 |
| WZ(2l2q) (Madgraph) | 3200975 | 2.29 | 1398.9 |
| ZZ(2l2q) (Madgraph) | 1936724 | 0.69 | 2799.2 |
| ZZ(4l) (Madgraph) | 4777861 | 0.17 | 27554.0 |
| tW-channel t (POWHEG) | 497656 | 11.10 | 44.8 |
| tW-channel $\bar{t}$ (POWHEG) | 493457 | 11.10 | 44.5 |

# 3 Physics Objects

## 3.1 Electron

The reconstructed electrons are from the "gsfElectron" electron collection. The cut-based electron identification (ID) is used and we choose the "medium" working point as default ID criteria, which includes Particle-Flow (PF) based/pileup (PU)-corrected isolation and photon conversion rejection. We also explore other electron IDs for cross checks such as "loose" or "tight" working points for the cut-based electron ID, or the MVA electron ID.

Only electrons meeting the following requirements are considered for further event candidate composition.

- $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ and $|\eta| < 2.4$
- Cut-based "medium" working point

The electrons are required to match to trigger objects if electrons are used to trigger an event. The matching is done by requiring the $\Delta R < 0.1$ ($\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$) between reconstructed electron and electron trigger object.

### 3.2 Muon

The reconstructed muons are from the "muons" muon collection. The 2012 POG-recommended "tight" muon selection criteria has been used as default muon ID, and the relative track-base isolation ($Iso_{track}$) with $Iso_{track}$ <0.1 is chosen as default working point, which is among the recommendations of muon POGs. Other isolation working point or PF-based isolation have been used for cross checks.

Only the muons meeting the following requirements are further considered in the event candidate composition.

- $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ and $|\eta| < 2.4$
- "tight" muon with $Iso_{track}$ <0.1

The muons are required to match to trigger objects if muons are used to trigger an event. The matching criteria is to require $\Delta R < 0.1$.

### 3.3 Jets

The "ak5PFJets" jet collection is used as the default jet collection. We apply the "ak5PFL1FastL2L3" jet energy correction (JEC) and "ak5PFL1FastL2L3Residual" for MC simulation and data, respectively. As shown in Section 2 the corresponding global tags are "START53_V27::All" and "FT_53_V21_AN6::All" for simulation and data.

We only use jets passing the following requirements.

- $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.4$
- "loose" jet ID and "medium" PU-jet ID

The implementation and further discussion of pu jetID can be found in,

https://twiki.cern.ch/twiki/bin/view/CMS/PileupJetID,

https://hypernews.cern.ch/HyperNews/CMS/get/met/333/1.html.

B-tagging has been used in this analysis, and the combined secondary vertex (CSV) btagger is used to identify b-jets. The default working point is "medium" criteria, where the cut value on the CSV btagger output is >0.679 and the rate of tagging light jet as b-jet is about 1-2%. Other working points are explored to cross check the results.

## 4 Corrections and Validations

We correct MC simulations for possible data/MC differences due to PU, lepton efficiencies, and b-tagging efficiencies if applicable. Additional corrections such as normalizing LO MADGRAPH DY+Jets to NNLO precision, tuning on dilepton $q_{\mathrm{T}}$ spectrum are implemented. Details of these corrections on MC simulations are described below.

Throughout this note, we define 9 primary signal/control regions based on flavor combination and mass of the two leptons, as shown in Table 3. Among these 9 signal/control regions, the 4 same-flavor off Z-pole regions are treated as our signal regions to search for potential "excess", and the same-flavor on Z-pole and the $e\mu$ channel are treated as control regions to validate MC simulations.

In the following discussions, the "inclusive dilepton samples" are often used to derive and validate corrections. The dilepton primary datasets are used. The selections of the inclusive

Table 3: The definition of various signal/control regions in dilepton channels.

| Lepton flavor | $50 < m_{ll} < 70\,\text{GeV}$ | $70 < m_{ll} < 105\,\text{GeV}$ | $105 < m_{ll}\,\text{GeV}$ |
|---|---|---|---|
| $ee$ | $ee$ below Z-pole | $ee$ on Z-pole | $ee$ above Z-pole |
| $e\mu$ | $e\mu$ below Z-pole | $e\mu$ on Z-pole | $e\mu$ above Z-pole |
| $\mu\mu$ | $\mu\mu$ below Z-pole | $\mu\mu$ on Z-pole | $\mu\mu$ above Z-pole |

dilepton events are identical to the main event selection with the only additional requirement of being charge opposite.

- Leading (1st) and next-to-leading (2nd) leptons $p_{\text{T}} > 20\,\text{GeV}$ and $> 15\,\text{GeV}$, $|\eta| < 2.4$
- Opposite signed (OS)
- Both leptons are trigger-matched: $\Delta R_{(l,HLT)} < 0.1$

## 4.1 Pileup

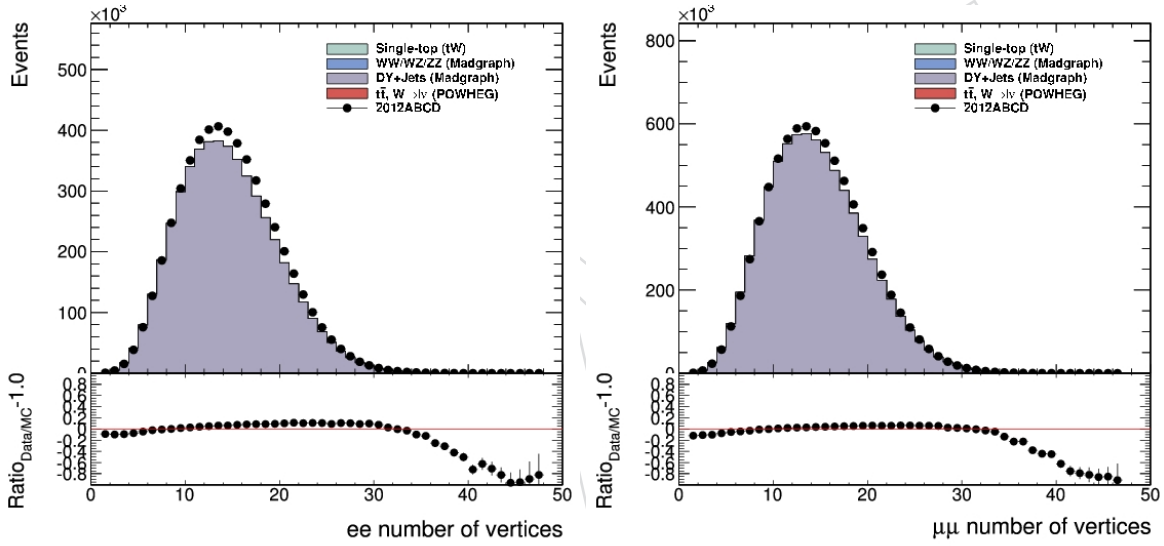


Figure 3: Number of reconstructed vertex for: Left) $ee$ channel and Right) $\mu\mu$ channel with $70 < m_{ll} < 105$ GeV. All corrections discussed in Section 4 have been applied in MC simulations. The MC simulations are normalized to data luminosity using cross section values in Table 2.

The MC simulations are produced with slightly different distribution of multiple interactions than what is expected in data. All MC simulations are reweighted to match to the expected PU distributions in data. We use the so-called 1-D pileup weighting. The expected PU distributions in data are calculated with"–inputLumiJSON" file,

/afs/cern.ch/cms/CAF/CMSCOMM/COMM_DQM/certification/Collisions12/8TeV/PileUp/pileup_latest.txt

The "–calcMode" is set to "true" and minbias cross section of 69.4 mb is used. Fig. 3 shows the number of reconstructed vertexes for inclusive same-flavor dilepton events on the Z-pole after PU weighting has been applied. The data are from "DoubleElectron" and "DoubleMuon" primary datasets. The reconstructed number of vertexes have good agreement between data and MC simulations after PU weighting.

## 4.2 Lepton Efficiencies

We determine the lepton efficiencies in 9 $p_T$ bins

15, 20.0, 25.0, 30, 35.0, 40, 45.0, 70, 100, 1000.0

and 15 $\eta$ bins

-2.4, -2.10, -1.85, -1.57, -1.44, -1.00, -0.60, -0.20, 0.20, 0.60, 1.00, 1.44, 1.57, 1.85, 2.10, 2.40

The binning is chosen to be identical between electron and muon. In addition we take into account the transition regions in both the electromagnetic calorimeter and the muon system.

The total lepton efficiency $\epsilon$ can be broken down into offline efficiency ($\epsilon_{off}$) and trigger efficiency ($\epsilon_{trg}$),

$$\epsilon^{e,\mu} = \epsilon^{e,\mu}_{off} \cdot \epsilon^{e,\mu}_{trg} \quad (1)$$

where $\epsilon^{e,\mu}_{off} = \epsilon^{e}_{reco} \cdot \epsilon^{e}_{ID}$ or $\epsilon^{\mu}_{sta} \cdot \epsilon^{\mu}_{ID} \cdot \epsilon^{\mu}_{iso}$ for electron and muon, respectively. The $\epsilon^{e}_{reco}$ and $\epsilon^{e}_{ID}$ are electron reconstruction and ID efficiencies. The $\epsilon^{\mu}_{sta}$, $\epsilon^{\mu}_{ID}$, and $\epsilon^{\mu}_{iso}$ are standalone muon reconstruction, muon ID, and muon isolation efficiencies. For double lepton trigger, $\epsilon^{e,\mu}_{trg}$ is defined to be one-leg efficiency.

All individual efficiencies are determined using the tag-and-probe (T&P) method with single lepton primary datasets and inclusive DY+Jets MC simulations. In each event, we only choose one T&P pair. For $\epsilon^{e}_{reco}$ and $\epsilon^{e}_{ID}$, we use supercluster and "gsfElectron" as probes, respectively. For the $\epsilon^{\mu}_{sta}$, $\epsilon^{\mu}_{ID}$, and $\epsilon^{\mu}_{iso}$, we use "generalTracks", "globalMuons", and "globalMuons" passing "tight" ID as probes. The trigger efficiency is calculated using leptons passing ID/isolation as probes.

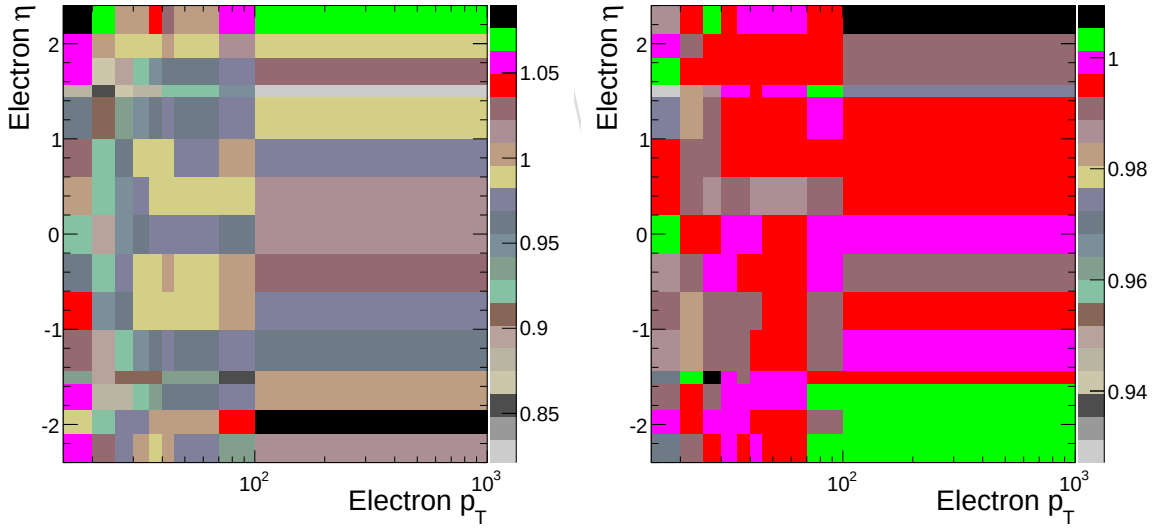


Figure 4: Scale factors for: Left) electron offline reconstruction ($\epsilon^{e}_{off}$), Right) one-leg trigger efficiency for double electron trigger ($\epsilon^{e}_{trg}$).

Fig. 4 and Fig. 5 show the data/MC scale factors determined using the T&P method for offline and trigger efficiency for electron and muon, respectively. All these scales and efficiency values are compared to the ones provided by e/gamma and muon POGs in common regions, and the agreement is at the level of 1-2%.

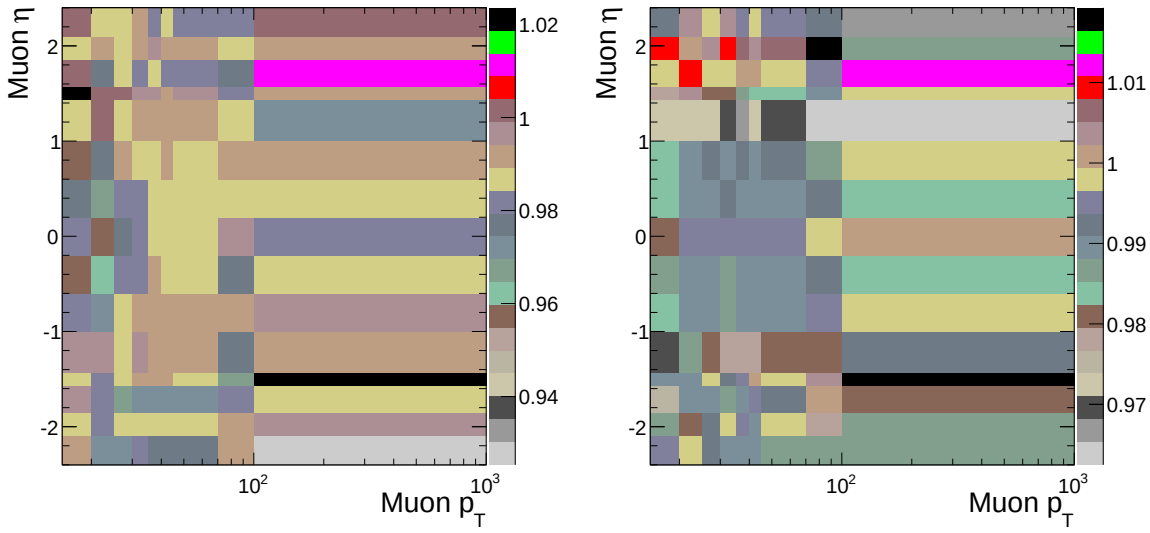


Figure 5: Scale factors for: Left) muon offline reconstruction ($\epsilon^{\mu}_{off}$), Right) one-leg trigger efficiency for double muon trigger ($\epsilon^{\mu}_{trg}$).

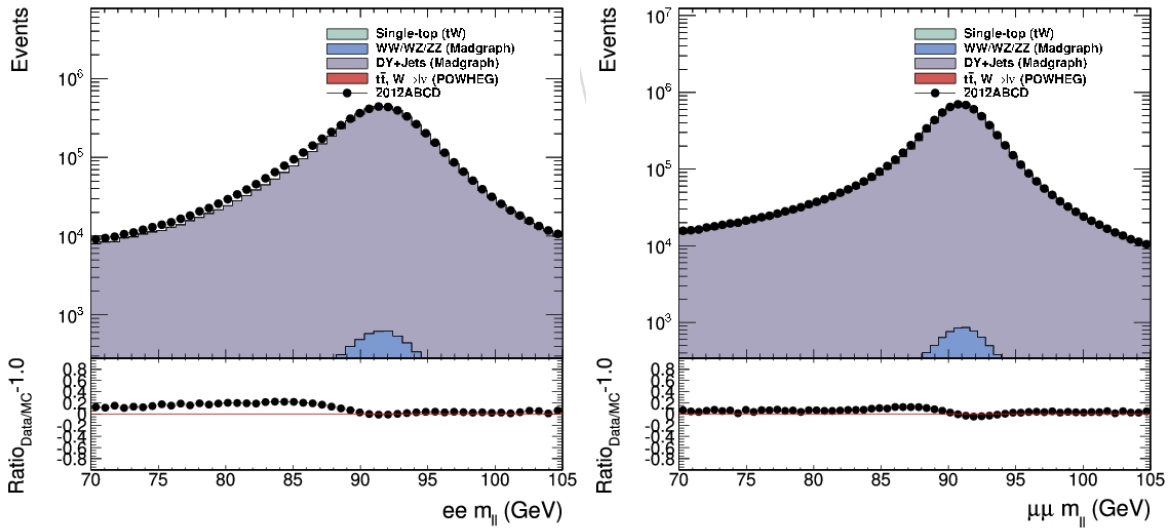


Figure 6: The dilepton invariant mass spectrum $m_{ll}$ for events on the Z-pole. Left) $ee$ channel and Right) $\mu\mu$ channel. All corrections discussed in Section 4 have been applied in MC simulations. The MC simulations are normalized to data luminosity.

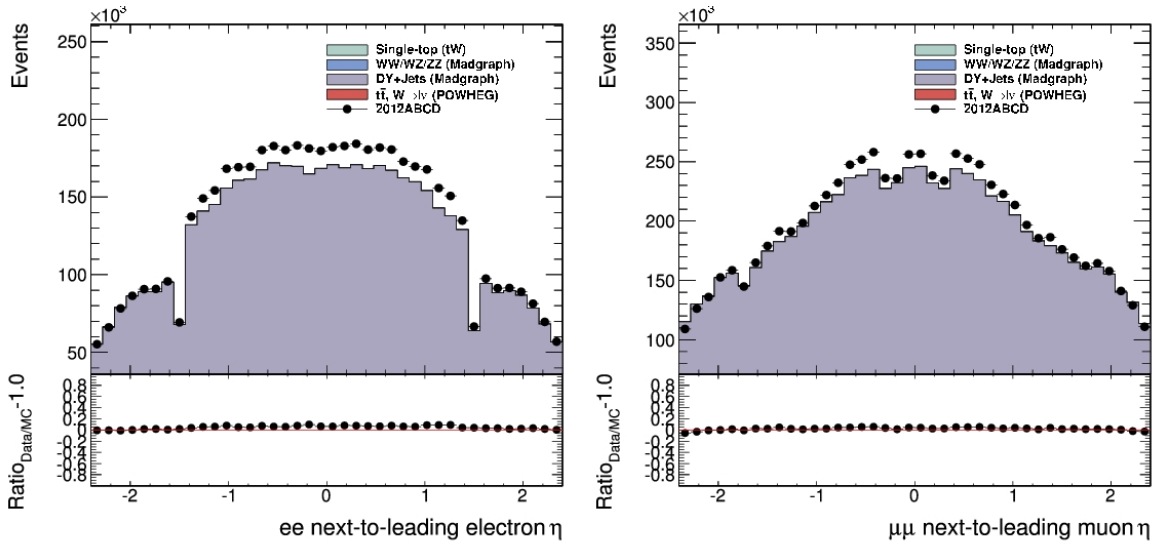


Figure 7: The tailing lepton $\eta$ distribution for events on the Z-pole. Left) *ee* channel and Right) $\mu\mu$ channel. All corrections discussed in Section 4 have been applied in MC simulations. The MC simulations are normalized to data luminosity.

Fig. 6 shows dilepton invariant mass spectrum $m_{ll}$ for events on the Z-pole for both *ee* and $\mu\mu$ channels. The $\eta$ distributions of the 2nd lepton is also shown in Fig. 7. In general good agreement between data and MC have been achieved. The data to total MC predictions are about 5.6% and 2.7% higher for *ee* and $\mu\mu$ channels, respectively.

## 4.3 B-tagging

When b-tagging is applied, we need correct for the data/MC differences in b-jet tagging efficiency or c/light/gluon mistag rates. To do this, we define an event-based probability, $\mathcal{P}(n^b; j_1, j_2, ..., j_N)$, which is the probability to observe $n^b$ b-tagged jets given a total of $N$ jets in the event with a tagging efficiency of $\epsilon_i$ (i=1, ..., N) for each jet

$$\begin{aligned}\mathcal{P}(n^b; j_1, j_2, ..., j_N) &= \Pi_{i=1...N}(1-\epsilon_i) \\ &\cdot(\Sigma_{(k_1\neq k_2\neq ...\neq k_{n^b})}\frac{\epsilon_{k_1}}{1-\epsilon_{k_1}}\cdot\frac{\epsilon_{k_2}}{1-\epsilon_{k_2}}\cdots\frac{\epsilon_{k_{n^b}}}{1-\epsilon_{k_{n^b}}})\end{aligned} \quad (2)$$

Following Eq. 2, for an event with 0 observed b-tagged jet, the above equation is simply as follows

$$\mathcal{P}(n^b = 0; j_1, j_2, ..., j_N) = \Pi_{i=1...N}(1-\epsilon_i) \quad (3)$$

To correct for data/MC differences in b-jet tagging efficiency or c or light/gluon mistag rates, we first calculate the $\mathcal{P}^{MC}(n^b; j_1, j_2, ..., j_N)$ using the MC truth b-jet tagging efficiency or c or light/gluon mistag rates. Then for each jet we correct the MC efficiency or mis-tag rates to what expected in data, using the scale factors provided by the Btagging POG [13]. We then use the corrected efficiency and mistag rates to calculate the $\mathcal{P}^{Data}(n^b; j_1, j_2, ..., j_N)$. The scale factor defined below can be used to weight every MC event to match the expected b-tagger performance in data. This whole procedure is applied to MC simulations only.

$$k_b = \frac{\mathcal{P}^{Data}(n^b; j_1, j_2, ..., j_N)}{\mathcal{P}^{MC}(n^b; j_1, j_2, ..., j_N)} \quad (4)$$

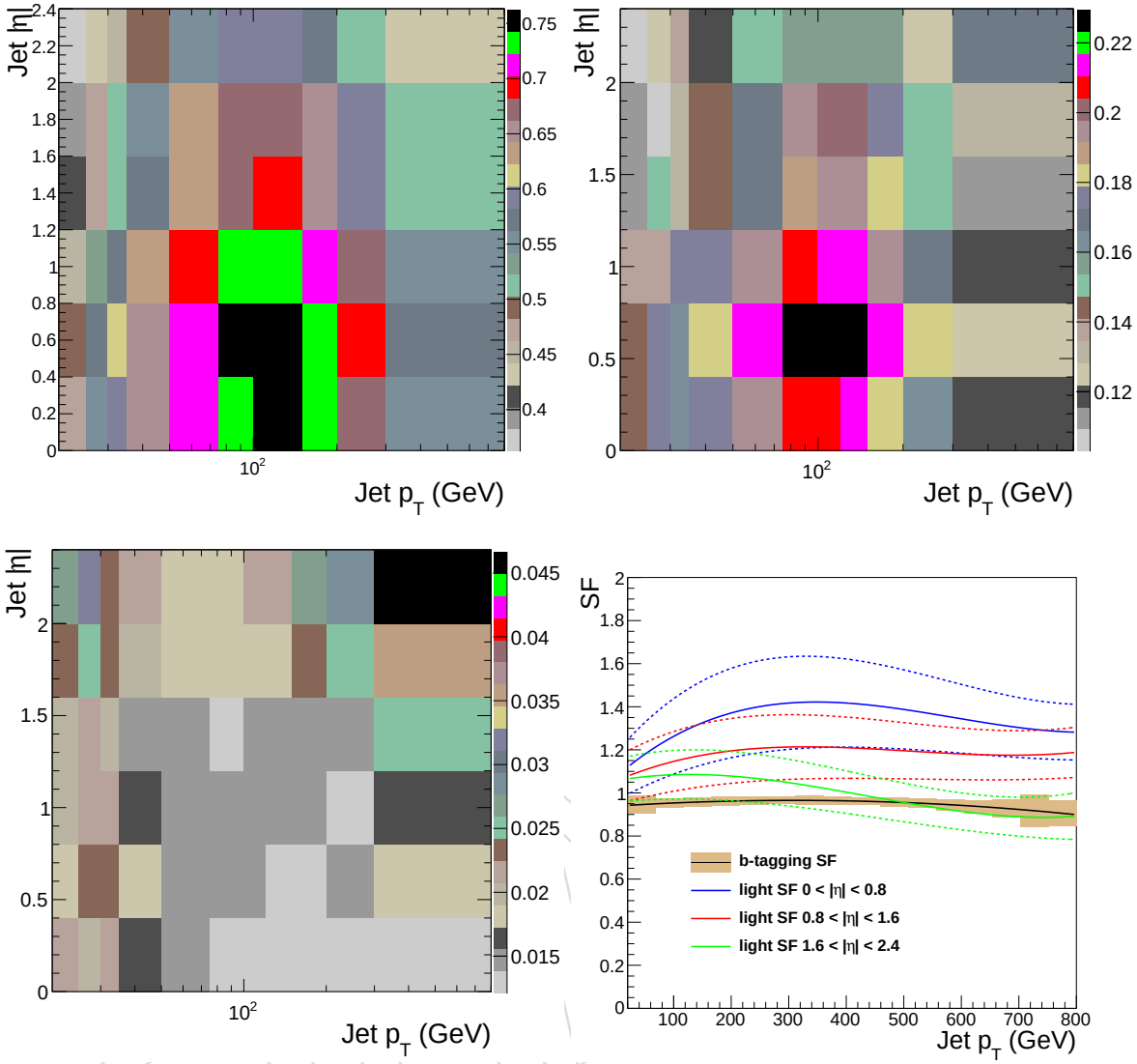


Figure 8: MC truth tagging efficiencies for Top-left) b-jet tagging efficiency Top-right) c mistag rate Bottom-left) gluon/light mist tag rate Bottom-right) Scale factors and uncertainties for b-jet tagging efficiency and c or light/gluon mistag rates. This is for the CSV "medium" btagger. All MC truth efficiencies are determined using $t\bar{t}$ MC simulations and the data/MC scale factors are from the btagging POG.

Fig 8 shows the MC truth b-jet tagging efficiency and c or gluon/light mistag rate for the CSV "medium" btagger. All these values are determined using the $t\bar{t}$ MC simulation in 11 $p_T$ bins

20, 25, 30, 35, 50, 75, 100, 120, 150, 200, 300, 800

and 6 $|\eta|$ bins

0, 0.4, 0.8, 1.2, 1.6, 2.0, 2.4

These values determined from MC simulations are combined with the scale factors for b-jet tagging efficiency or c or light/gluon mistag rates, which is also shown in Fig 8, to determine the scale factors to be applied in MC simulations.

## 4.4 Additional MC Tuning

### 4.4.1 The k-factor for DY+Jets Simulation

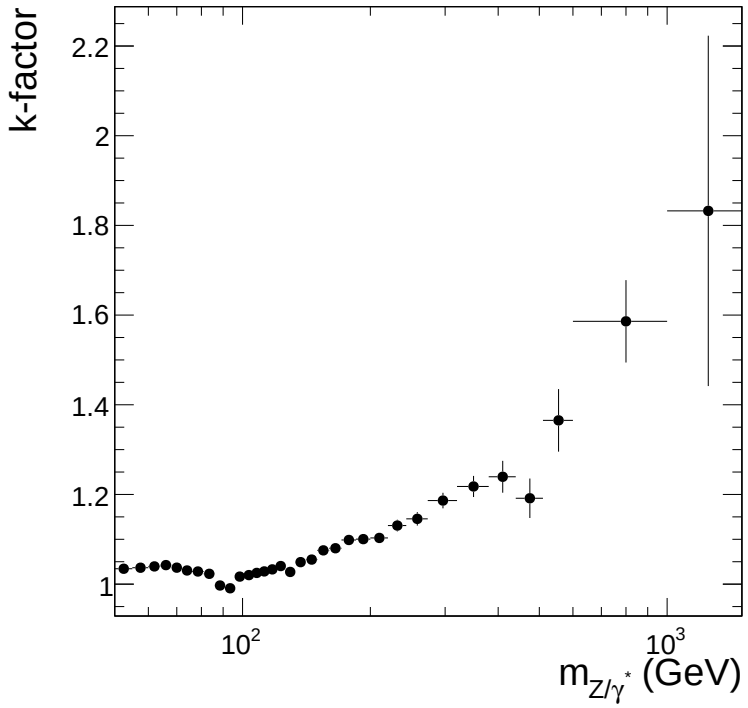


Figure 9: The mass-dependent k-factors $k_i$ to be applied to DY+Jets MC simulation. This correction is applied based on the generator level $Z/\gamma^*$ invariant mass, before the photon final-state radiation (FSR).

For the DY+Jets MC simulation, the cross sections predicted by MADGRAPH generator is of leading-order (LO) precision. We normalize the MADGRAPH DY+Jets total cross section to the one predicted by FEWZ 3.1 [14] at the next-to-next-to-leading order (NNLO) of $\alpha_S$, $1177.3^{+5.9}_{-3.6}(theory) \pm 38.8(PDF)$ pb. The DY line shape has been measured by CMS to very hihg precision [15] and in good agreement with predictions by FEWZ3.1. To have additional improvement of the modeling of the DY line-shape we calculate mass-dependent k-factors $k_i$ as below

$$k_i = \frac{\frac{1}{\sigma^{fewz}_{tot}} \cdot \sigma^{fewz}(m_i)}{\frac{1}{\sigma^{madgraph}_{tot}} \cdot \sigma^{madgraph}(m_i)} \quad (5)$$

where $\sigma_{m_i}$ and $\sigma_{tot}$ are the cross section in the $Z/\gamma^*$ mass bin $m_i$ and total cross section with $m_{Z/\gamma^*} > 50\,\text{GeV}$, respectively. The obtained k-factors are shown in Fig. 9. In general, the MADGRAPH MC simulation predicts a slightly narrower line shape around the Z-pole. Off the Z-pole, the k-factors are above unity. This correction is applied based on the generator level $Z/\gamma^*$ invariant mass, before the photon final-state radiation (FSR).

### 4.4.2 The Modeling of DY dilepton $q_T$ Spectrum

The $q_T$ spectrum of the Z boson has been measured by CMS with very high precision, and the comparison to the DY+Jets madgraph sample shows discrepancy [16]. Fig. 10 shows the $q_T$ spectrum of the inclusive $\mu\mu$ on Z-pole events. The $t\bar{t}$ background is subtracted using $e\mu$ events

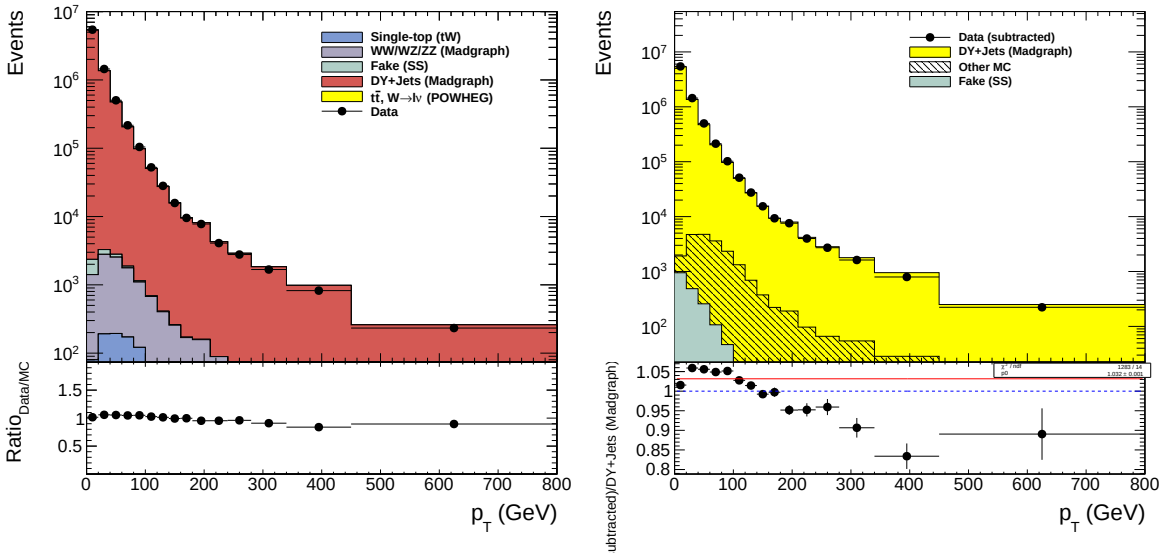


Figure 10: The $q_T$ spectrum of the $\mu\mu$ on Z-pole events. Left) data is compared to MC simulations, Right) background (all except DY+Jets) subtracted data compared to the DY+Jets MC simulation. The $t\bar{t}$ background is subtracted using $e\mu$ events in the same $m_{ll}$ mass region, after correcting for the electron and muon efficiency differences. All other corrections except the dilepton $q_T$ correction have been applied in MC simulations. The MC simulations are normalized to data luminosity.

in the same $m_{ll}$ mass region, after correcting for the electron and muon efficiency differences. The jet fake is determined using same sign (SS) $\mu\mu$ events and the contribution is negligible. The MADGRAPH MC simulation tends to produce slightly harder $q_T$ spectrum than what is observed in data. The difference derived from Fig. 10 is consistent to what has been measured [16]. This difference is used to correct for the DY+Jets MC simulation to improve the agreement of the lepton $p_T$ spectrum. This correction is applied based on the reconstructed dilepton $q_T$ values.

### 4.4.3 The Modeling of $t\bar{t}$ dilepton $q_T$ Spectrum

The $q_T$ of the reconstructed $t\bar{t}$ system tends to be softer in data than several MC simulations [17]. Instead of using the reconstructed $t\bar{t}$ system to explore this potential mismodeling, we use the $q_T$ of the dilepton system. Fig. 11 shows the $q_T$ spectrum of the inclusive $e\mu$ events above Z-pole. The "Mu-Electron" primary dataset is used. Similarly, the POWHEG MC simulations shows harder $q_T$ spectrum than what is observed in the data, and we correct POWHEG $t\bar{t}$ MC simulation for this difference. This is applied to $t\bar{t}$ MC simulation based on the reconstructed dilepton $q_T$.

## 4.5 Validation plots

Fig. 12 shows the inclusive dilepton invariant mass and $q_T$ spectrum after applying all the corrections. The whole invariant mass spectrum are in good agreement between data and MC simulations. The *ee* $q_T$ spectrum is not used to derived correction factors for DY+Jets MC simulation. However, the corrections derived using $\mu\mu$ events only are applied to *ee* events and good improvement for $q_T$ spectrum in *ee* channel has been achieved.

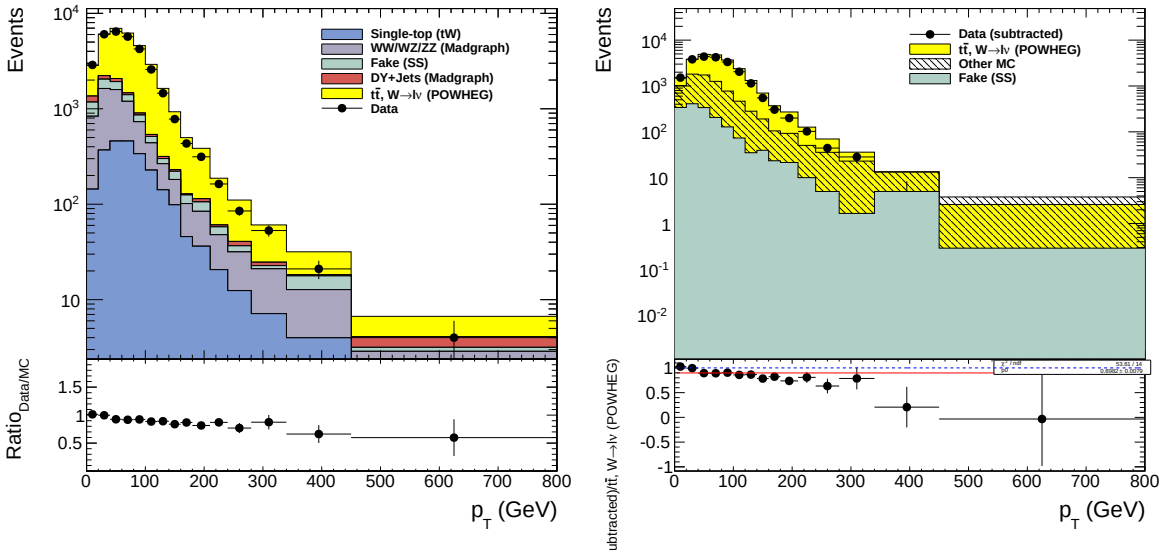


Figure 11: The $q_{\mathrm{T}}$ spectrum of the $e\mu$ events above Z-pole. Left) data is compared to MC simulations, Right) background (all except $\mathrm{t\bar{t}}$) subtracted data compared to the $\mathrm{t\bar{t}}$ MC simulation. The background subtraction is based on MC simulations. All other corrections except dilepton $q_{\mathrm{T}}$ weighting for $\mathrm{t\bar{t}}$ MC simulation have been applied. The MC simulations are normalized to data luminosity.

# 5 The Analysis Method

## 5.1 Event Selection

We select $lljj$ events using criteria described in Table 4. We only use certified objects as described in Section 3. For each event, we only consider all leptons and jets which pass the lepton ID and jet ID criteria and stay with the $p_{\mathrm{T}}$-$\eta$ acceptance for further event composition. After ID selections, all objects in each category such as electrons, muons, or jets are sorted according to the $p_{\mathrm{T}}$. The leading two leptons and leading two jets are used to reconstruct $lljj$ final state. We require the presence of the third jet in the sample, however, this 3rd jet is not included in calculating any 4-body observables. For jets passing ID criteria and $p_{\mathrm{T}} > 25\,\mathrm{GeV}$, $|\eta| < 2.4$ we count the number of jets ($n_b$) which pass the CSV "medium" working point. These events ("inclusive 3 jet/0 b-jet" sample) compose our primary data sample.

The "inclusive 3 jet/0 b-jet" data sample are finally divided into 9 regions as described in Table 3 based on lepton flavor composition and invariant mass ($m_{ll}$). The main result presented in this analysis note is based on this primary data sample. Additional control regions are also defined based on the number of the jets and b-jets, and these control regions are also examined. Studies of them will be included in an updated version this note.

## 5.2 The $m_{lljj}$ Mass Spectrum

Summary of $m_{lljj}$ mass spectrum in the 9 primary regions are shown in Fig. 13. All MC simulations are normalized using the cross sections shown in Table 2. The corrections to MC simulations discussed in Section 4 have been applied for MC simulations only. In general, below Z-pole the DY MC simulation predicts slightly lower yields than in data. This is partly due to the fact that a invariant mass cut on $m_{\mathrm{Z}/\gamma^*} > 50\,\mathrm{GeV}$ has been applied in the DY+Jets MC simulation. One thing worthy to point out the binning used in $m_{lljj}$ mass spectrum is 30 bins for a mass range of [50, 1850] GeV, which has been fixed since the 2011 analysis.

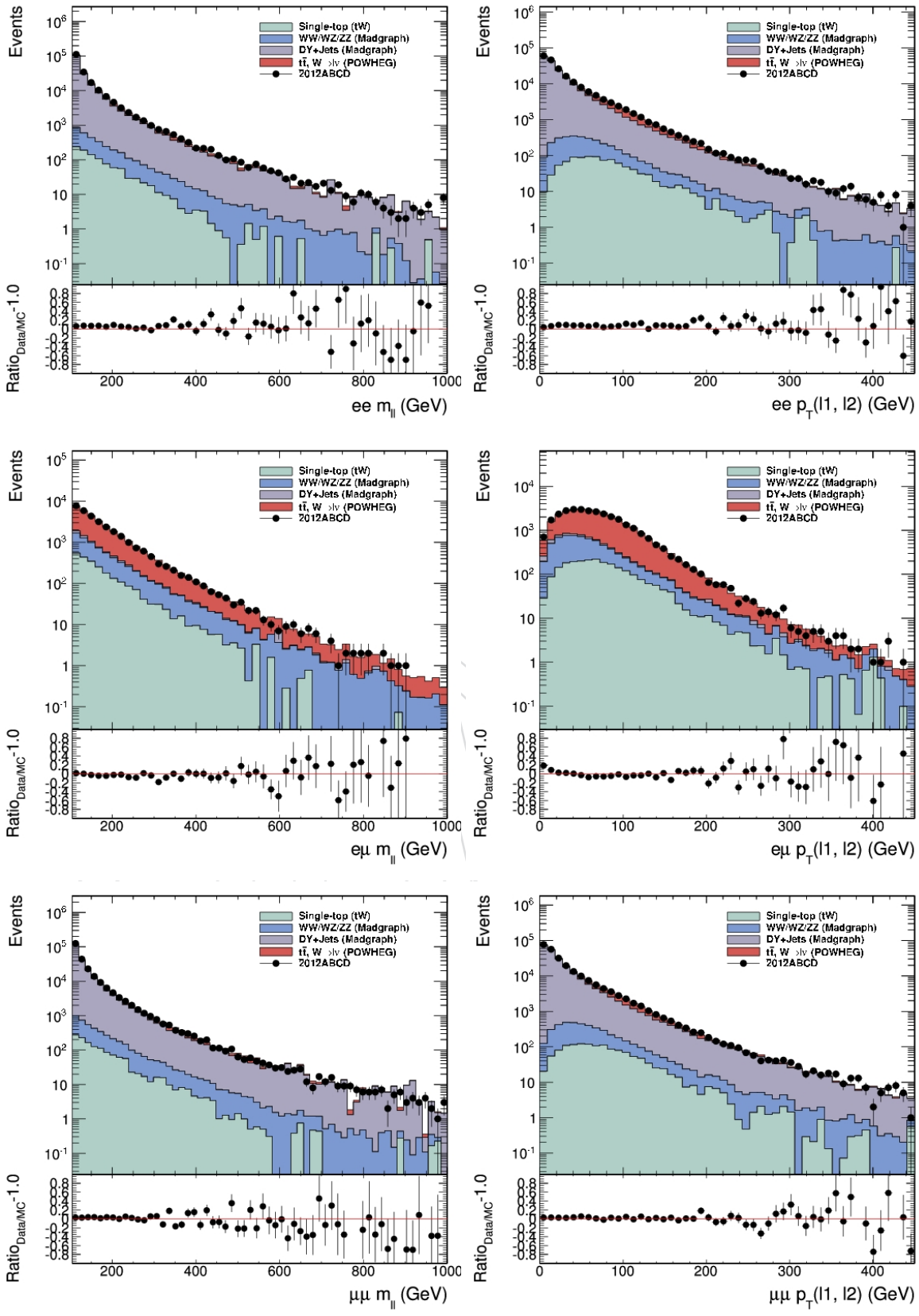


Figure 12: The inclusive dilepton invariant mass and $q_T$ spectrum for Top) $ee$, Middle) $e\mu$, and Bottom) $\mu\mu$ channels. All corrections discussed in Section 4 have been applied in MC simulations. The MC simulations are normalized to data luminosity using cross sections.

| Description | $ee$ | $e\mu$ | $\mu\mu$ |
|---|---|---|---|
| Leptons | cut-based "medium" gsfElectron, "tight" muon with $Iso_{track} < 0.1$<br>$p_{\mathrm{T}} > 15\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.4$ | | |
| | leading $p_{\mathrm{T}}^{e} > 20\,\mathrm{GeV}$<br>tailing $p_{\mathrm{T}}^{e} > 15\,\mathrm{GeV}$<br>electrons $\Delta R_{e,HLT} < 0.1$ | $p_{\mathrm{T}}^{\mu} > 20\,\mathrm{GeV}$<br>$p_{\mathrm{T}}^{e} > 15\,\mathrm{GeV}$<br>$\Delta R_{e/\mu,HLT} < 0.1$ | leading $p_{\mathrm{T}}^{\mu} > 20\,\mathrm{GeV}$<br>tailing $p_{\mathrm{T}}^{\mu} > 15\,\mathrm{GeV}$<br>muons $\Delta R_{\mu,HLT} < 0.1$ |
| Jets | $p_{\mathrm{T}} > 25\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.4$<br>"loose" jet ID/"medium" PU jetID/$\Delta R_{l,j} > 0.5$ (to both leptons)<br>1st $p_{\mathrm{T}} > 135\,\mathrm{GeV}$<br>2nd $p_{\mathrm{T}} > 30\,\mathrm{GeV}$<br>3rd $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ | | |
| Number of b-jets ($n_b$) | $n_b = 0$ | | |
| $m_{ll}$ | below Z-pole: $50 < m_{ll} < 70\,\mathrm{GeV}$<br>on Z-pole: $70 < m_{ll} < 105\,\mathrm{GeV}$<br>above Z-pole: $105 < m_{ll}\,\mathrm{GeV}$ | | |

Table 4: $ll + jj$ candidate selection criteria. The $\Delta R_{l,j}$ is the distance between the lepton and the selected jet.

In Fig. 13, across these 9 mass spectrum, in the $\mu\mu$ above Z-pole mass spectrum it appears to have a very distinguished Gaussian-like "excess" around 1 TeV, comparing to MC simulations. The "excess" is across 3-4 adjacent bins. The bin width is 60 GeV in all these figures. In the above Z-pole $ee$ and $e\mu$ channels, no indication of similar "excess" in the same mass regions. In the below Z-pole regions, the statistics is rather poor. In the following sections, we will perform statistical tests on these 9 distributions to quantify any local "excess". To avoid the turn-over region around 400-500 GeV, we restrict our tests to the mass range above 650 GeV. Fig. 14 shows the $m_{lljj}$ mass spectrum in $\mu\mu$ above Z-pole region, the same as the one in Fig. 13 with events in $m_{lljj} > 650\,\mathrm{GeV}$.

## 5.3 Evaluation of Significance

In this section, we will try to evaluate the significance of the "excess" in all 9 regions, including $\mu\mu$ above Z-pole channel. After all the event selections have been applied, the normalization of the MC simulations, which is based on the cross sections, tends to be slightly off, e.g. in Fig. 14 the MC simulations tends to over-estimate expected yields in data. We need rescale the MC normalizations.

For the two major SM processes: DY+Jets and $t\bar{t}$ backgrounds, we assume that there is no "excess" presented in data and take the following two steps to determine additional normalization factors

1. In each of the three $e\mu$ channel (above Z-pole, on Z-pole, and below Z-pole), we subtracted off the DY+Jets, single top and diboson background in data, the background subtracted data is compared to $t\bar{t}$ MC simulation to determine the corresponding normalization factor for $t\bar{t}$ MC simulation.

2. After applying the $t\bar{t}$ normalization factor derived from first step, in the $ee$ and $\mu\mu$ channels, we subtract off the $t\bar{t}$, single top and diboson background in data. The background subtracted data is compared to DY+Jets MC simulation to determine the normalization factor for DY+Jets background in each region.

After rescaling the major backgrounds, we scan the mass spectrum in the range of [700, 1440] GeV

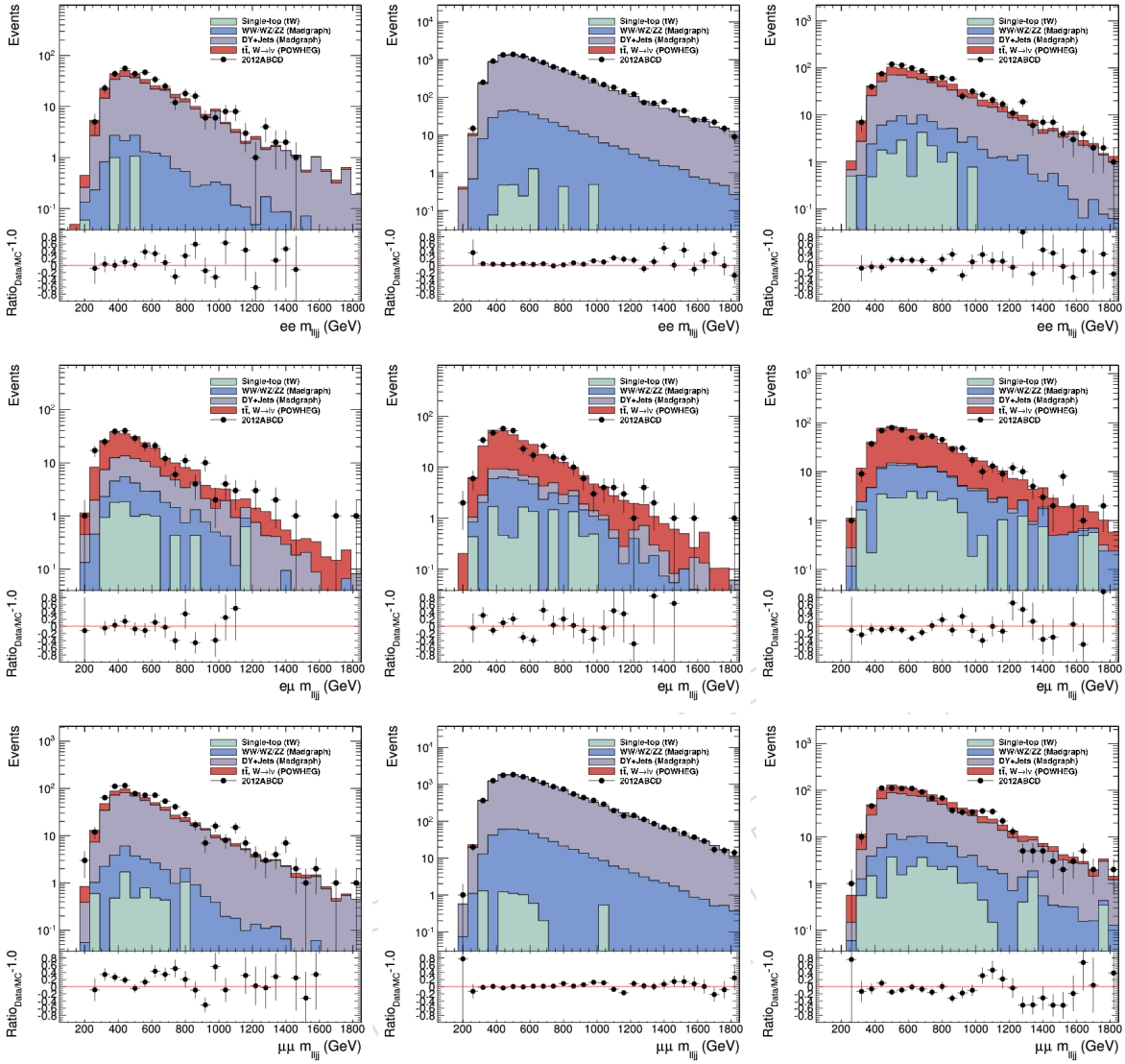


Figure 13: Summary of $m_{lljj}$ mass spectrum in 9 primary regions. Top) $ee$, Middle) $e\mu$, Bottom) $\mu\mu$. In each row, Left) below Z-pole, Middle) on Z-pole, Right) above Z-pole. The MC simulations are normalized to the data luminosity using cross section values in Table 2.

with a sliding mass window. In each mass window, we use MC simulations to estimate the known SM backgrounds, and a counting experiment is used to evaluate the probability of observe an "excess" in data. This is done using the CMS Higgs combination package [18]. We use the "ProfileLiklihood" method to evaluate the local p-value and significance. The details of deriving the background normalization and performing the statistic tests are discussed in the following sections.

### 5.3.1 Normalization of the $t\bar{t}$ Background

The additional normalization factor for $t\bar{t}$ MC simulation is determined as follows using $e\mu$ primary dataset.

1. Fig. 15 shows the $m_{lljj}$ distributions in mass range of [650, 1850] for $e\mu$ channels above Z-pole, on Z-pole, and below Z-pole. The DY+Jets, single top and diboson background

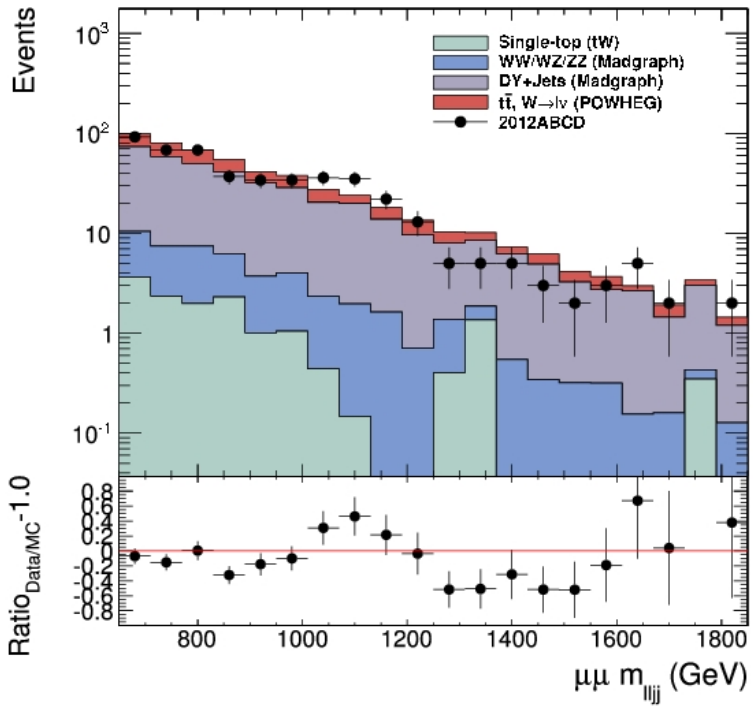


Figure 14: The $m_{lljj}$ mass spectrum in $\mu\mu$ above Z-pole region with $m_{lljj} > 650\,\text{GeV}$. The MC simulations are normalized to the data luminosity using cross section values in Table 2.

are estimated based on MC simulations and are subtracted from the data, the background subtracted data is compared to $t\bar{t}$ MC simulation and the bin-to-bin ratios are shown. To remove the empty bins, we rebin the histogram by 2 iteratively until there is no empty bins in the raw data distributions. The additional scale factors for $t\bar{t}$ MC simulation are derived by fitting the ratio plot with a constant using a chisquare method, which is shown in Fig. 15. The fitted constants and statistical errors are summarized in Table 5. These are the additional scale factors to be applied for $t\bar{t}$ MC simulations in $e\mu$ channels.

2. In this step, we derive the scale factors for $t\bar{t}$ MC simulations in both $ee$ and $\mu\mu$ channels. Mainly due to different $p_T$ dependence of lepton efficiencies in low $p_T$ regions, the acceptance for $t\bar{t}$ events are slightly different in $ee$ and $\mu\mu$ channels than in $e\mu$ channels. This can bring slightly difference for the data/MC scale factors. We repeat the same procedure using same $e\mu$ channel data and MC simulations as the first step. However, additional two corrections are applied before performing the background subtractions and fitting for the scale factors. We use the $ee$ channel to illustrate the procedure.

   - For each event in $e\mu$ MC simulation, we treat the muon as an electron and apply the the efficiency scale factors between the electron data efficiency and muon MC efficiency on the MC sample. For $e\mu$ data, we treat the muon as an electron and apply the efficiency scale factor between the electron data efficiency and muon data efficiency. After doing so, we treat the $e\mu$ $t\bar{t}$ events as $\mu\mu$ $t\bar{t}$ events.

     To derive the scale factors in $\mu\mu$ channel, we simply treat the electron as muon and apply corresponding scale factors.
   - We correct for the acceptance difference due to the muon $p_T > 20\,\text{GeV}$ and

electron $p_T > 15\,GeV$ selections in $e\mu$ events. Due to this particular $e\mu$ $p_T$ combination, events with muon $p_T$ below 20 but above 15 GeV, and electron $p_T$ above 20 GeV are not kept in the selected $e\mu$ sample. This is done by simply applying a weight factor of 2 for events with electron $p_T$ within [15, 20] GeV. This is done for both $e\mu$ channel data and MC simulations.

After above two additional corrections, we subtract single top, diboson, and DY+Jets background in data and fit the ratios between data/$t\bar{t}$ MC to derive the scale factors. The obtained scale factors are also summarized in Table 5.

During above procedure, for MC simulations in $e\mu$ channels we do not apply the trigger matching. Trigger efficiencies for electron and muon are applied to MC simulation and this is equivalent to apply the trigger match and applying data/MC scale factors for triggers.

| Lepton flavor | below Z-pole | on Z-pole | above Z-pole |
|---|---|---|---|
| $ee$ | $0.817 \pm 0.256$ | $1.18 \pm 0.168$ | $0.935 \pm 0.0777$ |
| $e\mu$ | $0.916 \pm 0.263$ | $1.18 \pm 0.167$ | $0.947 \pm 0.0778$ |
| $\mu\mu$ | $0.964 \pm 0.277$ | $1.16 \pm 0.167$ | $0.958 \pm 0.0788$ |

Table 5: Scale factors to be used to rescale $t\bar{t}$ MC simulation in each of the 9 signal/control regions.

Fig 16 shows comparisons of the $t\bar{t}$ $m_{lljj}$ spectrum in MC simulations between $ee/\mu\mu$ channels and the ones obtained using $e\mu$ channel after applying above two corrections. All $t\bar{t}$ MC simulations are normalized to data luminosity. The ratios between the $ee/\mu\mu$ mass spectrum to the one in $e\mu$ channels are very close 0.5. The luminosity in these three double lepton datasets are almost same. This illustrates that after applying above two corrections the acceptance in $e\mu$ channel $t\bar{t}$ events is well matched to the one in $ee/\mu\mu$ channels.

### 5.3.2 Normalization of DY+Jets Background

For the $ee/\mu\mu$ channels, where DY+Jets background dominates, we determine additional normalization factors. The procedure is pretty much similar as what has been done to derive the scale factors for $t\bar{t}$ backgrounds. We first subtracted off the single top, diboson, and $t\bar{t}$ background from data. Here the $t\bar{t}$ scale factors from Table 5 have been applied. The background subtracted data is compared to DY+Jets MC simulation and the data/MC scale factors are derived by fitted to the ratio distributions. Fig. 17 shows the results and the scale factors for DY+Jets in $ee/\mu\mu$ are summarized in Table 6.

| Lepton flavor | below Z-pole | on Z-pole | above Z-pole |
|---|---|---|---|
| $ee$ | $1.15 \pm 0.129$ | $1.06 \pm 0.018$ | $1.09 \pm 0.0808$ |
| $e\mu$ | - | - | - |
| $\mu\mu$ | $1.24 \pm 0.105$ | $1.03 \pm 0.0155$ | $0.813 \pm 0.0656$ |

Table 6: Scale factors to be used to rescale DY MC simulation in $ee$ and $\mu\mu$ channels.

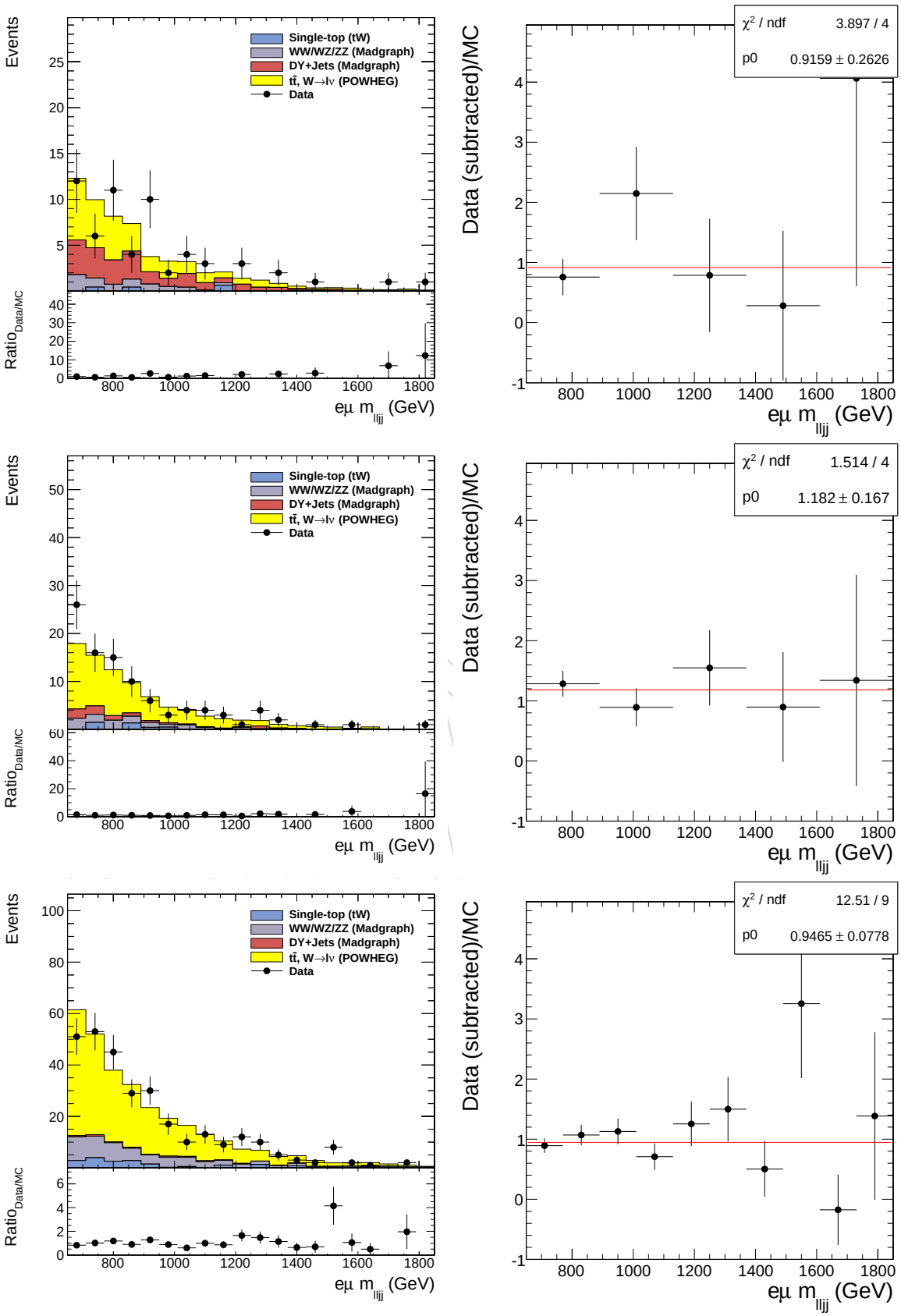


Figure 15: The $m_{lljj}$ mass spectrum in $e\mu$ channels for Top) below Z-pole, Middle) on Z-pole, Bottom) above Z-pole. For the stacked plots on the left, the data is normalized to data luminosity using cross sections. On the right, ratios between the background subtracted data over $t\bar{t}$ MC simulation are fitted with a constant (see text for more discussions).

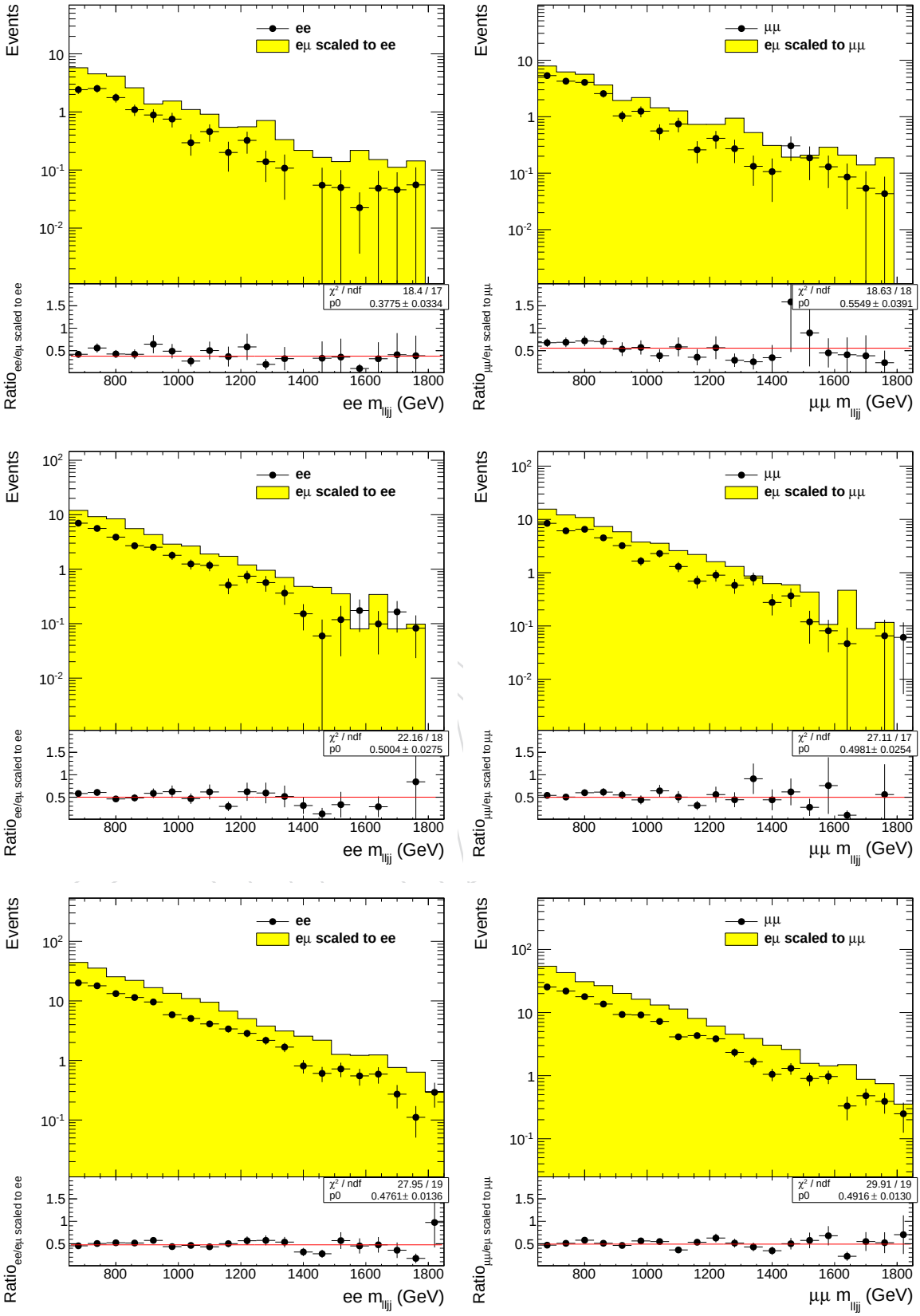


Figure 16: The $m_{lljj}$ mass spectrum in $t\bar{t}$ MC simulations for Top) below Z-pole, Middle) on Z-pole, Bottom) above Z-pole. The left ones are for $ee$ channel, and the right ones are for $\mu\mu$ channels. The $ee/\mu\mu$ channels are compared to the one in $e\mu$ channels after correcting for electron/muon efficiency difference and acceptance. (see text for more discussions).

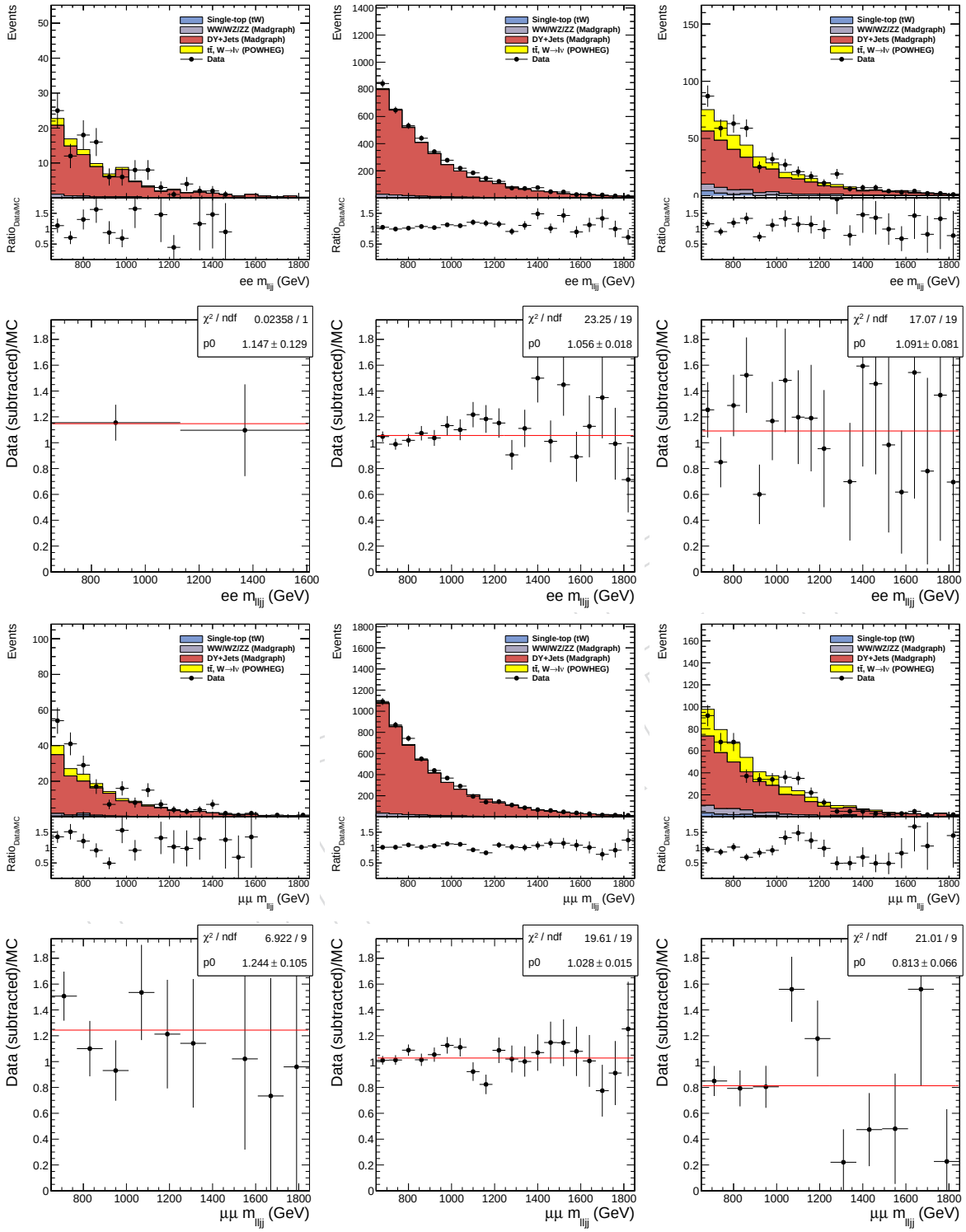


Figure 17: The $m_{lljj}$ distributions and background subtracted data compared to DY+Jets MC simulation for $ee/\mu\mu$ channels. The top two rows are for $ee$ and the bottom two rows are for $\mu\mu$ channels. From left to right are for below Z-pole, on Z-pole, and above Z-pole. The $t\bar{t}$ normalization has been rescaled using scale factors from Table 5.

### 5.3.3 The Statistical Method

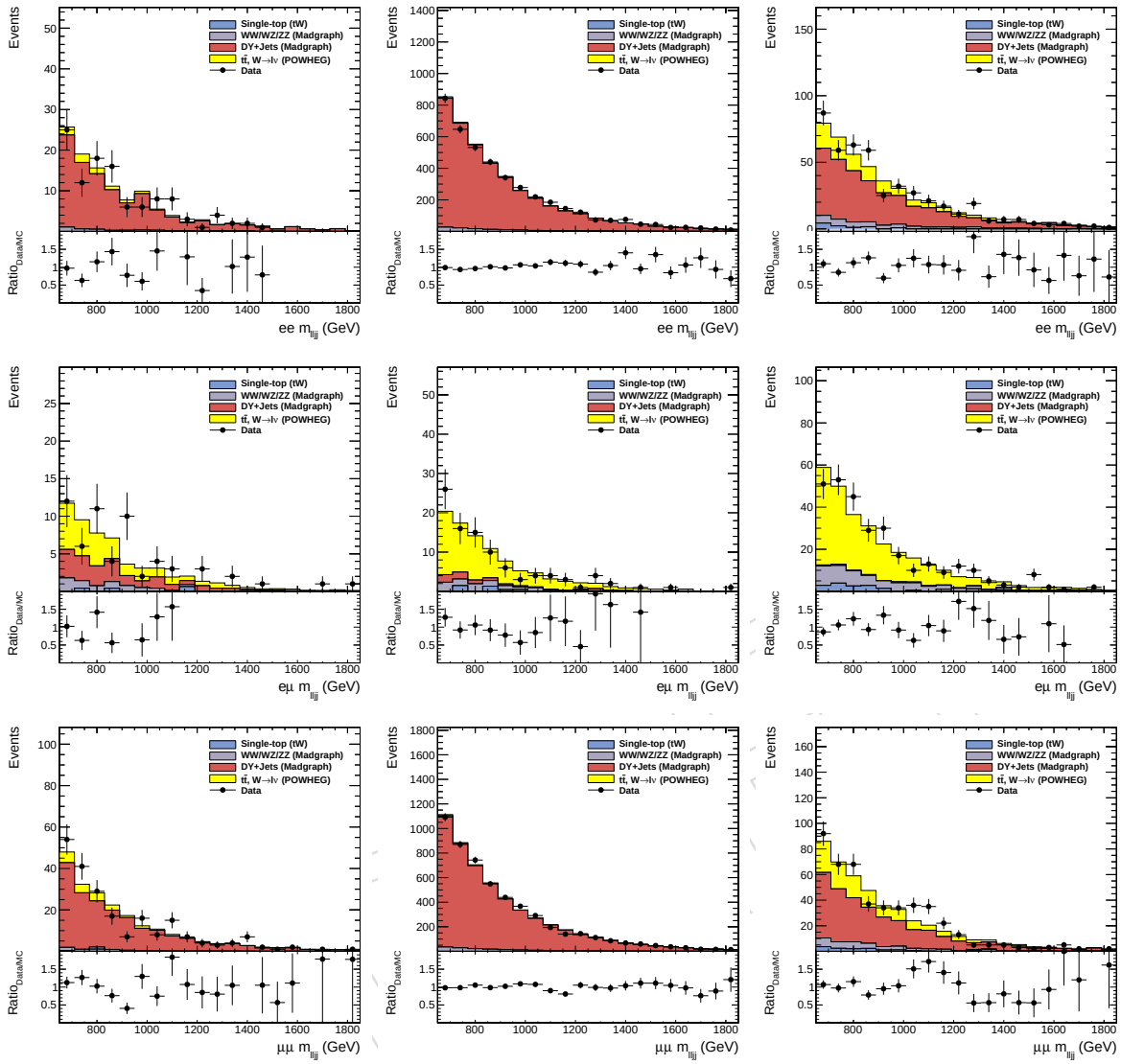


Figure 18: Summary of $m_{lljj}$ mass spectrum in 9 primary regions. Top) $ee$, Middle) $e\mu$, Bottom) $\mu\mu$. In each row, Left) below Z-pole, Middle) on Z-pole, Right) above Z-pole. The $t\bar{t}$ and DY+Jets MC simulations have been rescaled using scale factors in Table 5 and Table 6.

Fig. 18 shows the $m_{lljj}$ spectrum for $ee$ and $\mu\mu$ channels after applying scale factors for $t\bar{t}$ and DY+Jets MC simulations shown in Table 5 Table 6. A closer side-by-side comparison between $ee$ and $\mu\mu$ above Z-pole samples is shown in Fig 19.

We scan all 9 mass spectrum with a sliding mass window for any local deviation("excess") and determine the probability of any "excess" due to fluctuation of SM background. This is done using the CMS Higgs combination package [18] and a simple counting experiment is performed. The known SM background contributions are based on MC simulations. We use the "ProfileLikelihood" method in our statistic tests.

The width of the sliding mass window could be tricky given that we do not have a particular signal model. The right-handed W is less likely to be a potential explanation of the "excess" in $\mu\mu$ channel. However, the right-handed W can decay into the same final state as we are

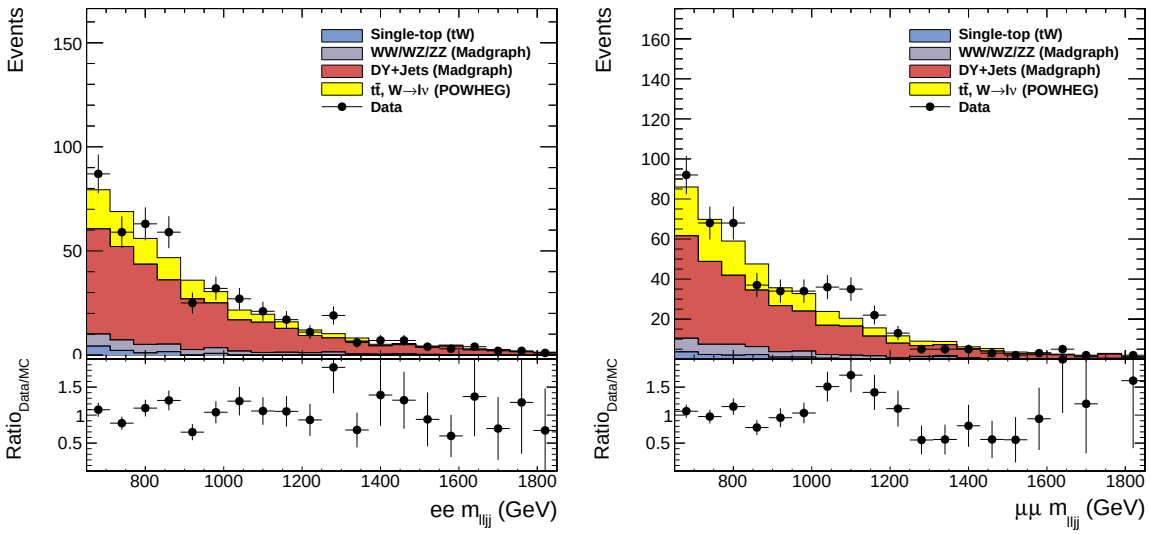


Figure 19: Summary of $m_{lljj}$ mass spectrum for Left) $ee$ and Right) $\mu\mu$ above Z-pole data samples. The $t\bar{t}$ and DY+Jets MC simulations have been rescaled using scale factors in Table 5 and Table 6.

exploring. The intrinsic width of the right-handed W is rather small, at the order of 10-20 GeV assuming the mass of right-handed W is about 1 TeV. Due to these two reasons, it could be an ideal candidate to provide some insights to what kind of detector resolution we would expect if the "excess" were due to the decay of a heavy object.

We perform the toy MC studies by using a parametrized background shape to generate background events and inject "signal" events using the MC simulation of a right-handed W mass of 1000 GeV and the heavy neutrino mass of 500 GeV. The details are discussed in Appendix A.2. In each toy, we fit the dataset with same background shape with a Gaussian as signal. The fitted relative Gaussian width is plotted against the fitted Gaussian width as shown in Fig. 20 for both a floating background shape and a fixed background shape. The fits with a floating background shape show largely bias in estimating the signal yield and the Gaussian peak due to radiation tails in right handed W signal shape and rather low statistics encountered in our case. By fixing the background shape to what is used to generate the background events, the pull of signal yield is much more improved. The fitted relative Gaussian widthes in both cases are about 10%. We simply take a 10% sliding mass window ([0.9m, 1.1m]) for our statistic tests.

# 6 Systematic Uncertainties

The following systematic uncertainties are considered at this moment. The dominant ones are due to the normalization of DY+Jets and $t\bar{t}$ backgrounds. This section is expected to be refined in further studies.

## 6.1 Luminosity

The single top and diboson backgrounds are normalized to data luminosity using the production cross sections in Table 2. We assign 2.6% uncertainty for the estimated single top and diboson background. These are completely correlated between single top and diboson background.

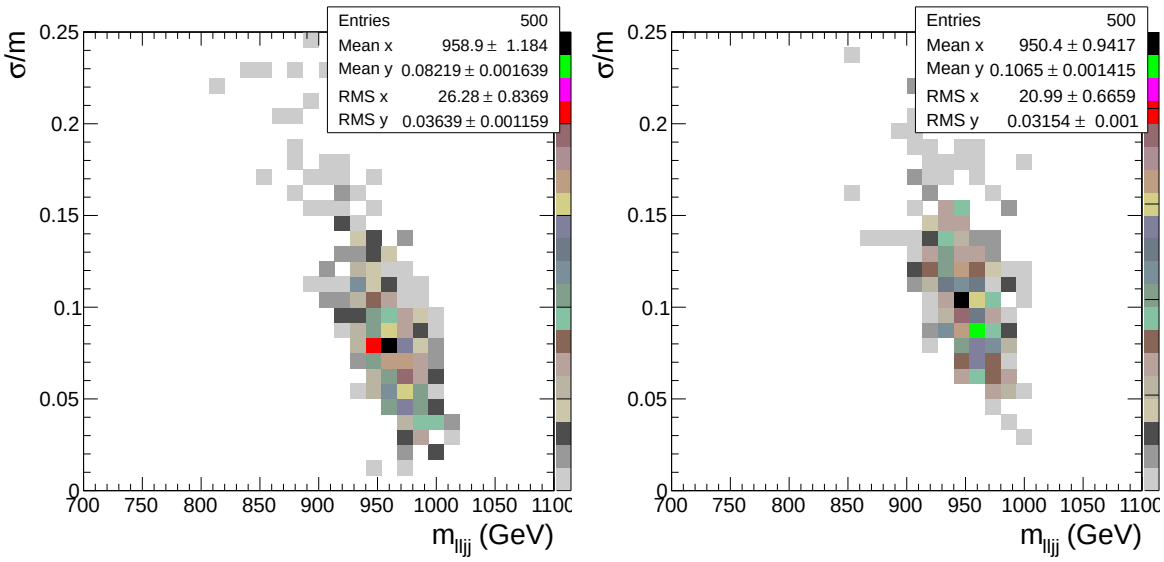


Figure 20: The fitted relative Gaussian width is plotted against the fitted Gaussian width for Left) floating background shape, Right) fixed background shape.

## 6.2 Pile up

work in progress

## 6.3 Production Cross Sections for Single Top and Diboson

Additionally, we assign 6.9% uncertainty on the production cross section of single top, which includes PDF uncertainty and scale uncertainties. For the diboson, we assign a correlated 20% uncertainty in each of WW, WZ, ZZ production cross section. This is the difference between the CMS measurement of WW production and the theoretical predictions at the NNLO.

## 6.4 Production Cross Section for Drell-Yan

For $e\mu$ channel, the DY+Jets background are normalized using inclusive DY+Jets production cross section. We take a total of 3.3% uncertainty on the cross section, which takes into account the PDF uncertainty and scale variations.

## 6.5 The $t\bar{t}$ Normalization

We take the statistical uncertainties of the scale factors for the $t\bar{t}$ background in each region as systematic uncertainty. These are shown in Table 5. These uncertainties includes both statistical uncertainties in data and MC simulations. This is the second largest systematic uncertainty.

## 6.6 The DY+Jets Normalization

For $ee$ and $\mu\mu$ channels, we take the statistical uncertainties of the scale factors for the DY+Jets background in each region as systematic uncertainty. These are shown in Table 5. These uncertainties includes both statistical uncertainties in data and MC simulations. This is the dominant systematic uncertainty.

# 7 Final Results

## 7.1 Final Results

Fig. 21 shows the local p-value and the corresponding significance as a function of the center of the sliding mass window for the $e\mu$ channels. The systematic uncertainties discussed in previous section have been included. Across the whole mass range, no significant local "excess" is observed in each of three dilepton mass regions.

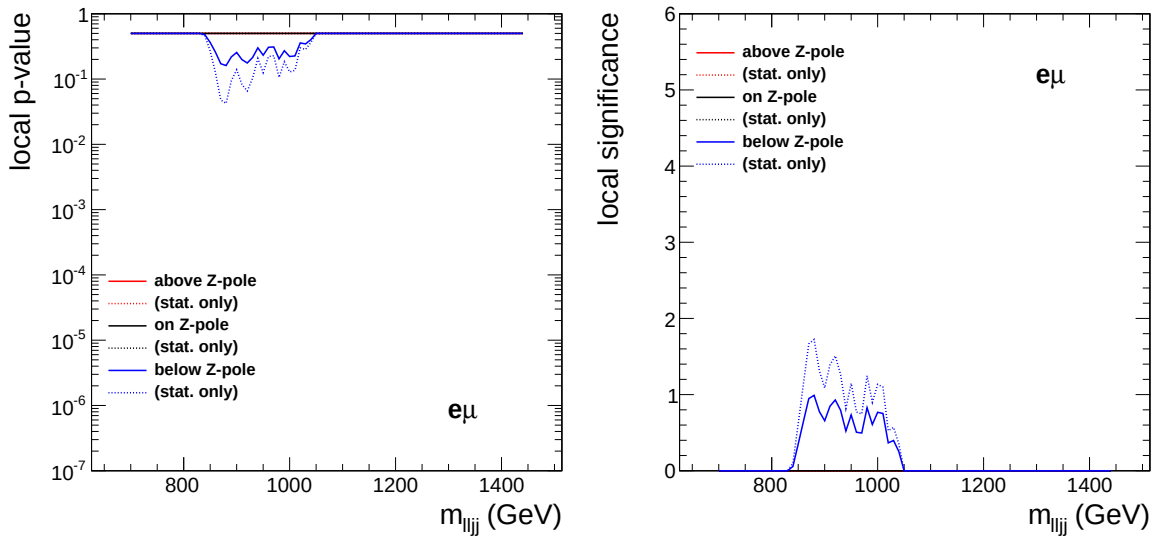


Figure 21: Left) the local p-value and Right) the local significance as a function of the center of the sliding mass window for the $e\mu$ different dilepton mass regions.

Fig. 21 shows the local p-value and the corresponding significance as a function of the center of the sliding mass window for the $ee$ and $\mu\mu$channels. For each channel, the mass spectrum in three different dilepton mass regions are shown. The systematic uncertainties discussed in previous section have been included. In the $ee$ channels, there is no significant local "excess" observed in all three different regions. However, in the $\mu\mu$ above Z-pole region, we observe an excess around 1 TeV mass with a local significance of 3.8 $\sigma$ after considering systematic uncertainties. Without considering systematical uncertainties, the local significance is 4.1 $\sigma$. Below the Z-pole, slight excess is also observed in the similar $m_{lljj}$ region, with maximum local significance of 1.6 $\sigma$. The local significance at mass point of 1090 GeV, which is the place that we observe the maximum local significance for above Z-pole sample, is about 1.3$\sigma$ with systematical uncertainty.

## 7.2 Background-subtracted Results

To examine closely the "excess" events, we subtract the expected SM backgrounds and compare kinematics distributions of the "excess" events with the SM background events. We choose the events in the mass window [981, 1199] GeV, where the scan of the sliding mass window gives the maximum statistical significance. This is the so-called "signal" region. We also choose a lower mass window [750, 870] GeV, where no "excess" of data over SM MC simulations is evident, to test how well the SM MC simulations simulate the kinematic distributions of known SM processes. This is our "lower sideband". The convention we take here is that leptons and jets are sorted in $p_{\mathrm{T}}$ and labeled as $l1$, $l2$ and $j1$, $j2$, $j3$, $j4$, respectively.

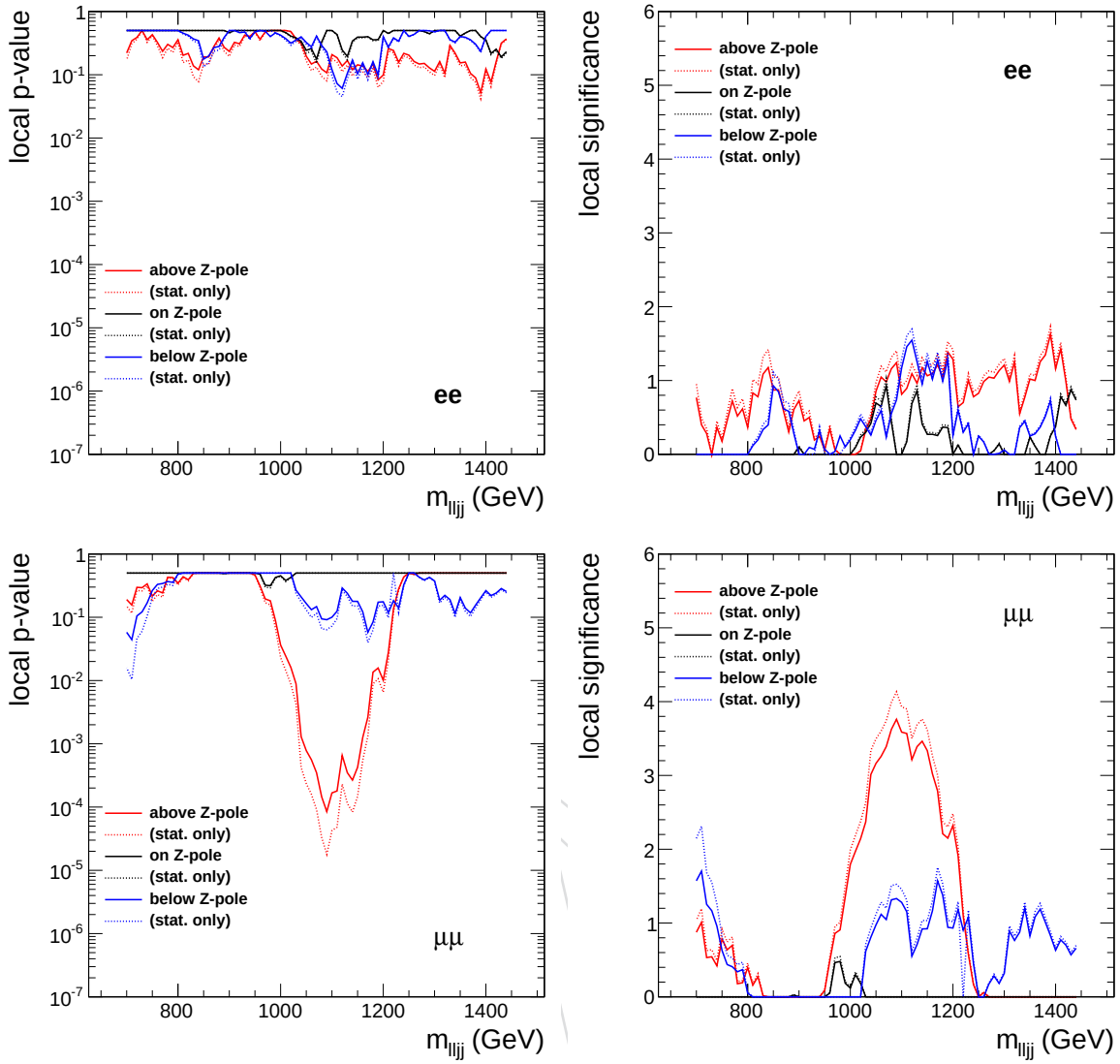


Figure 22: Left) the local p-value and Right) the local significance as a function of the center of the sliding mass window for the $ee/\mu\mu$ channels in different dilepton mass regions. The $ee$ channel is shown on the top and the $\mu\mu$ channel is shown in the bottom.

Fig. 23 shows 3-body invariant mass distributions $m_{l1l2j2}$ for "lower sideband" and "signal" region. The MC $t\bar{t}$ and DY+Jets contributions have been renormalized according to scale factors in Table 5 and Table 6. In the "lower sideband", the SM MC simulations can describe the kinematics of data reasonablely well.

Fig. 24 shows the background subtracted distribution for the events in "signal" region. The kinematic cuts for leading and 2nd muon are 20 and 15 GeV, respectively. The $p_{\mathrm{T}}$ cut on the 2nd jet is 30 GeV. From the "lower sideband", the MC can reliably predict the kinematic edge of the SM DY+Jets and $t\bar{t}$ events. However, for the "excess" events in signal region, the 3-body invariant mass tends to have a kinematic edge a couple of hundred GeV higher than what SM MC simulations predicts. Many other background subtracted distributions are shown in Appendix E.

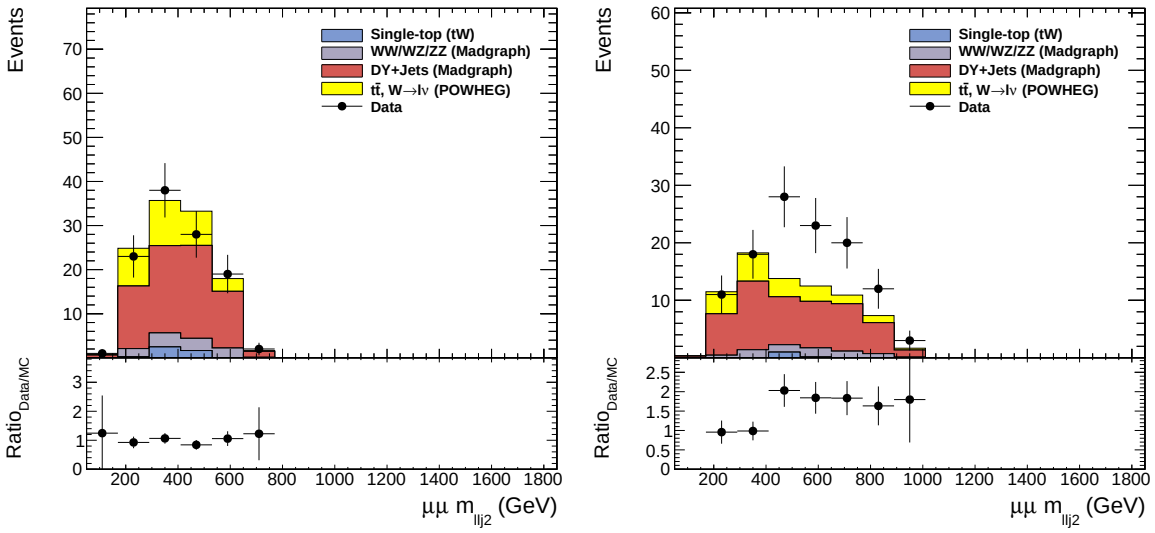


Figure 23: The 3-body invariant mass distributions $m_{l1l2j2}$ for Left) "lower sideband" and Right) "signal" region. The MC $t\bar{t}$ and DY+Jets contributions have been renormalized according to scale factors in Table 5 and Table 6.

# 8 Summary and Conclusion

Motivated by a potential "excess" spotted in 2011 CMS data we have systematically examined the $ll + jj$ mass spectrum using 2012 CMS data. This involves analyzing of 9 different mass regions defined by electron and muon flavor combination and mass range: above Z-pole, on Z-pole, and below Z-pole. After preliminary event selections, we scan the $m_{lljj}$ mass spectrum from [700, 1440] GeV with a 10% ([0.9m, 1.1m]) sliding mass window, using MC simulations to estimate known SM backgrounds. No significant local "excess" above SM expectation has been observed in all mass regions in $ee$ and $e\mu$ channels. In $\mu\mu$ above Z-pole region, we observe an excess around 1 TeV mass with a local significance of 3.7 $\sigma$ after considering systematic uncertainties. Without considering systematical uncertainties, the local significance is 4.1 $\sigma$. Below the Z-pole, slight excess is also observed in the similar $m_{lljj}$ region, with maximum local significance of 1.6 $\sigma$ after considering systematic uncertainties. The local significance at mass point of 1090 GeV, which is the place that we observe the maximum local significance for above Z-pole sample, is about 1.3$\sigma$ after considering systematical uncertainty.

In summary, we have found evidence of an "excess" over the expected SM production in $m_{lljj}$ mass spectrum at around 1 TeV. The local significance in one of our search regions is about 3.7$\sigma$ after considering systematic uncertainties. The significance of this "excess" is about 4.1$\sigma$ with statistical uncertainties only. Similar "excess" is not presented in the same kinematic regions of $ee$ or $e\mu$ channels. Additional cross checks and re-analyzing of the 2011 data is still ongoing and will be included shortly.

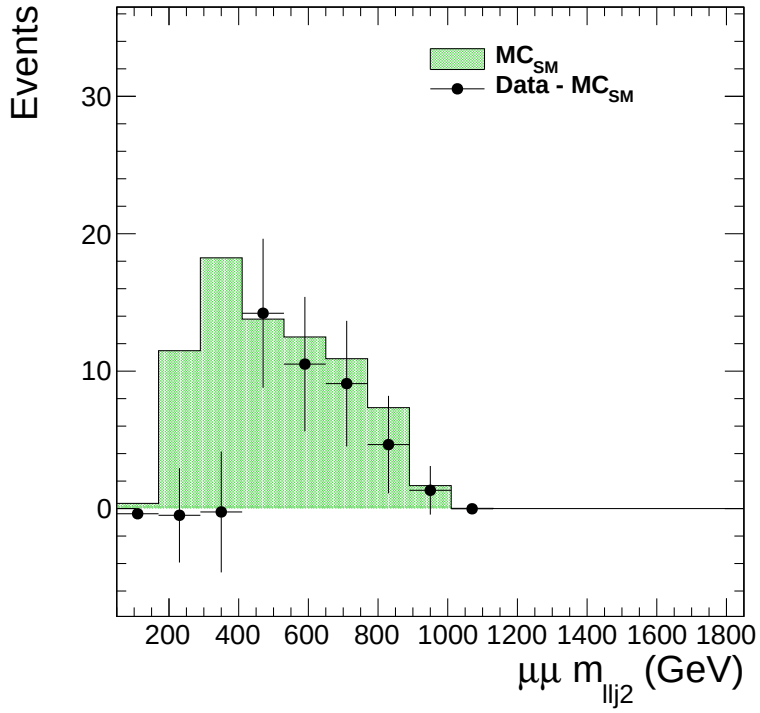


Figure 24: The background subtracted 3-body invariant mass distributions $m_{l1l2j2}$ for "excess" events in the "signal" region.

# A Cross check of $ll + jj$ mass peak between 2011 and 2012 datasets

## A.1 Comparison of fitted mass peak

Fig. 25 shows fits to the same range of mass spectrum ([350, 1850] GeV) using same signal and background functions for 2011A, 2011B, and 2012 datasets. The functions used here are a Gaussian signal function and a background function defined to be

$$f(x) = (1 - \frac{m}{\sqrt{s}})^{a} \cdot (\frac{m}{\sqrt{s}})^{-b-cln(\frac{m}{\sqrt{s}})} \tag{6}$$

where $s$ is the center-of-mass energy of pp collisions and $a$, $b$, $c$ are free parameters determined by the data itself. This is the background function used in CDF/CMS dijet searches. The Gaussian signal shape for fit to Run2011B dataset, including both mean and width, has been fixed to the one obtained from the fit to Run2011A dataset. This is due to the fact that it is not possible to fit for a Gaussian signal with both floating mean and width by Run2011B data alone. Same fit has been performed to the mass spectrum obtained using the first 1.1 $\text{fb}^{-1}$ luminosity of 2011A, where we have first spotted a possible "excess". This is shown in Fig. 26. The selection used in making this mass spectrum is different from others as the selection was intended to study other physics processes.

The fitted mass peaks in different datasets are summarized in Fig. 27. Note that the 1.1 $\text{fb}^{-1}$ of 2011A data has been included in the full 2011A dataset. The fitted mass peak shows a difference of 42 GeV between 2011A and 2012 datasets, where a clear peak can be identified. Given the two datasets are independent, the total error on the mass difference is 32.8 GeV

$$\Delta m = 42.0 \pm 32.8\,\text{GeV} \tag{7}$$

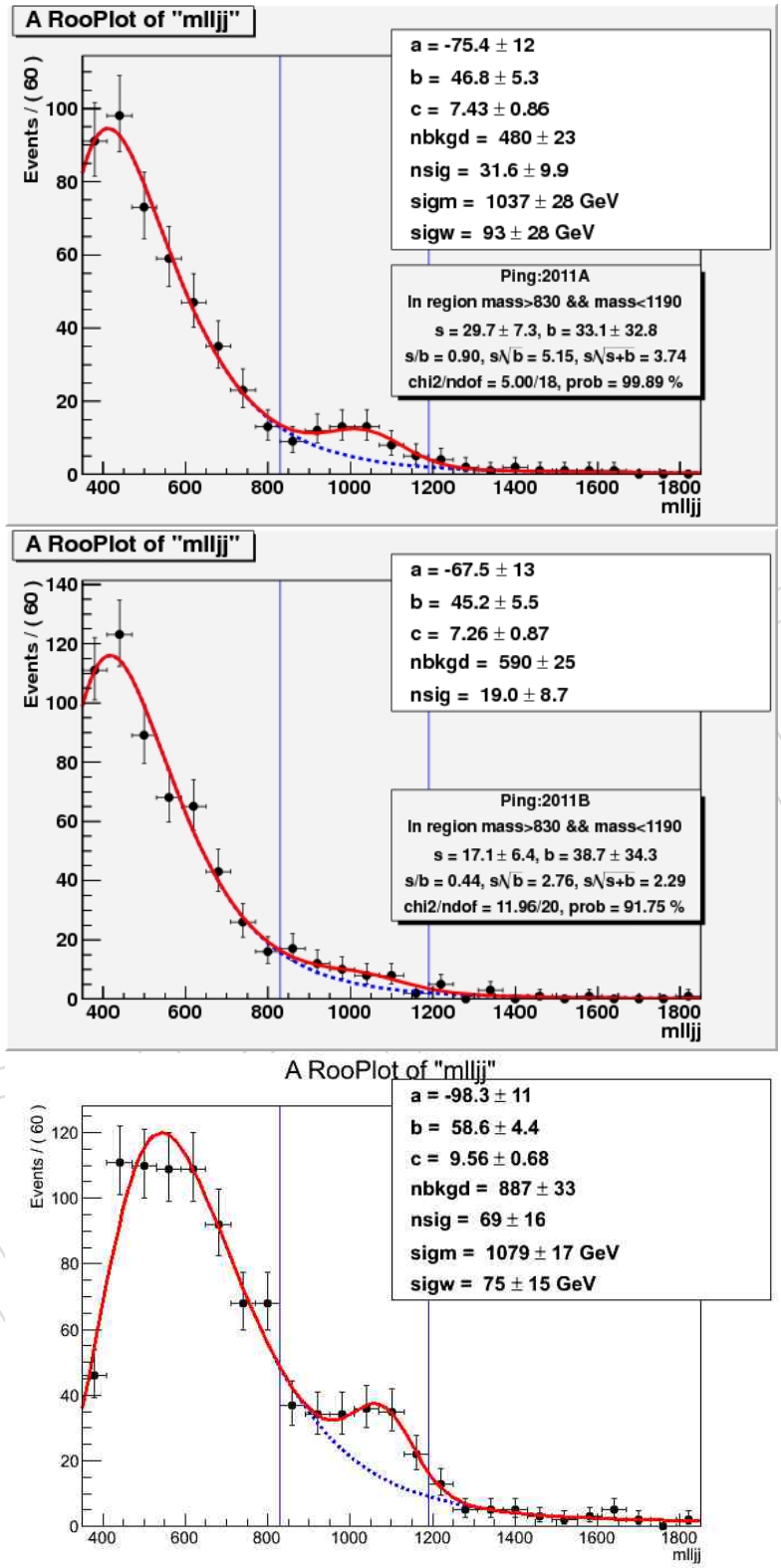


Figure 25: Signal + background fits to the mass spectrum in [350, 1850] GeV using same functions for 2011A, 2011B, and 2012 datasets. Note that different selections have been applied between 2011 and 2012 analyses, details of these difference are discussed in this analysis note and Ref [1, 3]. The fits to 2011A and 2011B are the ones from the cross check note [3]. The Gaussian signal shape for fit to Run2011B dataset, including both mean and width, has been fixed to the one obtained from the fit to Run2011A dataset.

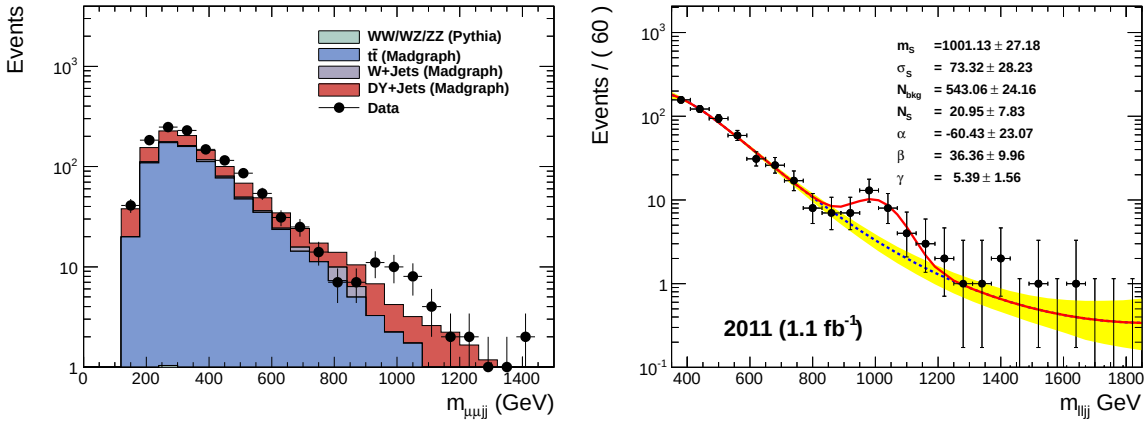


Figure 26: Left) Mass spectrum for the first 1.1 fb$^{-1}$ of 2011A dataset, which is the identical plot as shown in Fig. 1, Right) Signal + background fit to the mass spectrum in [350, 1850] GeV using same parametrization as in Fig. 25. The parameters $\alpha$, $\beta$, and $\gamma$ shown here are parameters $a$, $b$, and $c$ defined in Eq. 6, respectively.

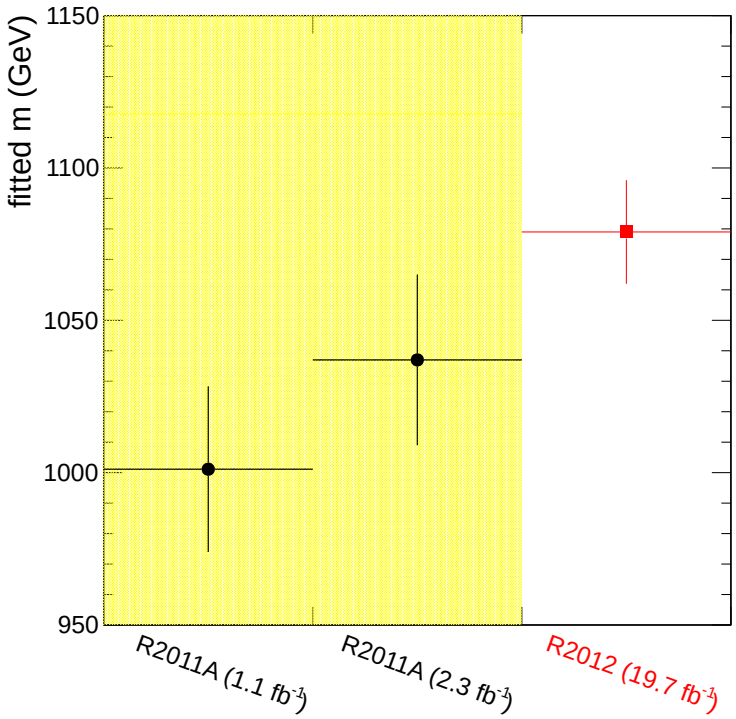


Figure 27: Summary of the fitted mass peak in different data subsets, as shown in Fig. 25 and Fig. 26. The Run2011B result is not included here since it is not possible to fit for a Gaussian peak by itself. The data sample used in fitting for the "Run2011A(1.1 fb$^{-1}$)" result is included in the data sample used in fitting for the "Run2011A(2.3 fb$^{-1}$)" result.

## A.2 Toy MC study with the fitting technique

This analysis is not targeted for any specific signal model. For 2011 analysis, we have relied on a fitting method to extract the yield of "excess" events. The background is parametrized by an empirical function defined in Eq. 6. Given the lack of a signal model, we simply treat the "excess" as a Gaussian and let the data determine the mean and width. One potential drawback to this simplified signal model is that non-Gaussian tails are presented for a high mass resonance with jets in the final state. In addition, the sample size of selected events are small, and the tail of background distribution might not be well constrained by the data itself.

We perform toy MC studies to understand potential bias in this fitting method. For this purpose, we take the right-handed W signal MC simulation as illustration as it decays into the same final state, however this is not our signal model. We take the MC simulation with right-handed W mass of 1000 GeV, and the heavy neutrino mass of 500 GeV. Same selection as 2011 analysis has been applied to this MC simulation, then events are randomly chosen from this MC sample for each toy. The background events are generated using the fitted function from the fit to Run2011A dataset. For each toy, both number of background events ($N_{bkg}$) and number of right-handed W signal events ($N_{sig}$) are sampled from Poisson distribution with mean of 1920 and 160, respectively. These are chosen to be roughly 4 times of the fitted "signal" and "background" events for 2011A so that most of toys can converge into a good fitted result. A total of 500 toys have been generated, and the fitted results are summarized in Fig. 28. The pull of the fitted signal yields shows that this fitting method has large bias for the yield and could potentially under-estimate the uncertainty of fitted signal yield. This is largely due to the strong correlation between the fitted signal and background yields, which is also shown in Fig. 28. The fitted Gaussian mean tends to have relatively large spread, which also correlated with the fitted width.

# B PU-dependence of W mass peak in MC simulation and $t\bar{t}$ data

We systematically validate the JEC in different run periods, different detector $\eta$ regions, and its residual dependence to PU using MC simulations and $t\bar{t}$ data.

Fig. 29 shows the reconstructed W mass using two leading jets in the WZ diboson MC simulation, which is generated with PYTHIA event generator. The events are selected using

- Leading and next-to-leading muon $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $> 15\,\mathrm{GeV}$, $|\eta| < 2.4$
- Muon charges are oppositely signed
- $60 < m_{\mu\mu} < 120\,\mathrm{GeV}$
- $n_{jets} \geq 2$ with jet $p_{\mathrm{T}} > 30\,\mathrm{GeV}$, $|\eta| < 2.4$
- Reconstructed dijet invariant mass, $60 < m_{jj} < 120\,\mathrm{GeV}$

With the selected MC events, we calculate the mean of $m_{jj}/m_W^{PDG} - 1$ as a function of number of vertexes in the event, which is shown in Fig. 29. The mean has a residual dependence on the number of vertexes. The two distributions are fitted with a straight line, and the slope is $0.0036 \pm 0.00038$ and $0.0015 \pm 0.00019$ for 7 TeV and 8 TeV MC simulations, respectively. As a reference, the pileup distributions in 2011A, 2011B, and 2012 are shown. The mean PUs are 6.2, 11.1, 19.9 for 2011A, 2011B, and 2012 datasets, respectively.

We also check the PU-dependence of reconstructed W mass in $t\bar{t}$ MC simulation and data. The selection of "muon+$\geq$4 jets" sample and reconstruction the W mass peak is as below

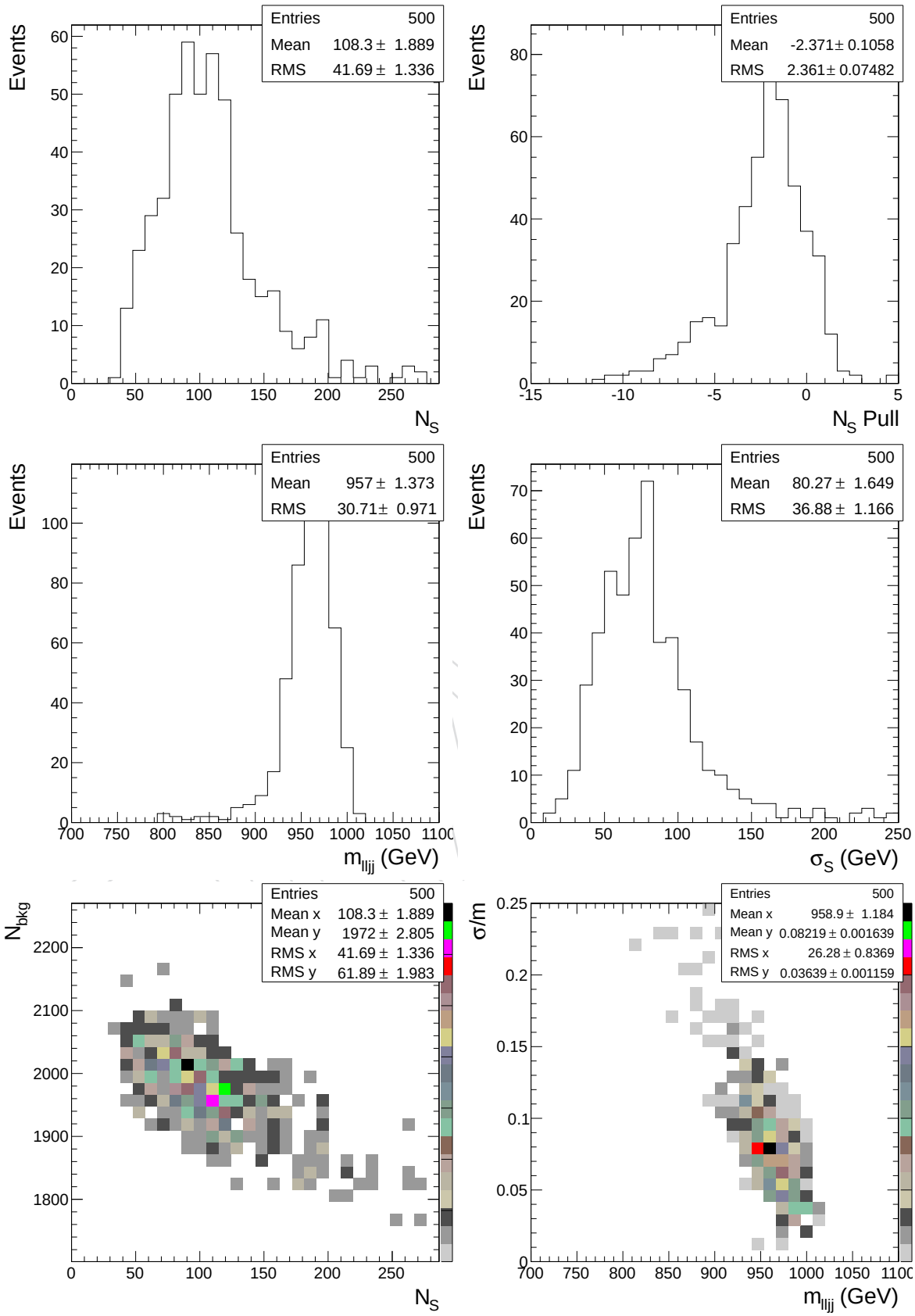


Figure 28: Top-left) fitted signal events, Top-right) pull distribution, Middle-left) fitted mean of Gaussian, Middle-right) fitted width of Gaussian, Bottom-left) fitted signal and background events, Bottom-right) the fitted mean and relative width.

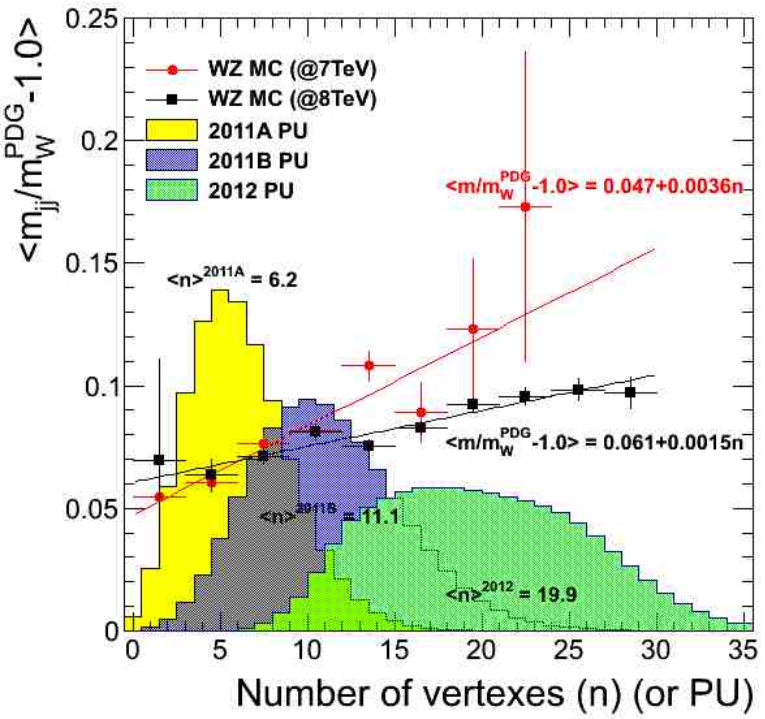


Figure 29: Mean of $m_{jj}/m_W^{PDG} - 1$ as a function of number of vertexes in the event. The two distributions are fitted with a straight line and the slope is $0.0036 \pm 0.00038$ (Red) and $0.0015 \pm 0.00019$ (Black) for 7 TeV and 8 TeV MC simulations, respectively. As a reference, the pileup distributions in 2011A, 2011B, and 2012 are shown. The mean PUs are 6.2, 11.1, 19.9 for 2011A, 2011B, and 2012, respectively.

- Leading muon $p_T > 25\,\text{GeV}$, $|\eta| < 2.4$ passing "tight" muon ID and $Iso_{track} < 0.1$
- Leading muon is trigger matched to HLT_IsoMu24* with $\Delta R(\mu, HLT) < 0.1$
- No additional "medium" cut-based electron or "tight" isolated muon with $p_T > 15\,\text{GeV}$, $|\eta| < 2.4$
- Only jets passing "loose" jetID and "medium" PU jetID with $\Delta R(\mu, jet) > 0.5$ are considered
- At least 2 CSV "medium" b-tagged jets with $p_T > 25\,\text{GeV}$, $|\eta| < 2.4$
- At least 2 light jets failed CSV "medium" working point with $p_T > 25\,\text{GeV}$, $|\eta| < 2.4$

The two leading light jets are used to reconstruct the W mass peak as shown in Fig. 30. Corrections due to PU, lepton efficiencies and b-tagging scale factors have been applied to MC simulations only. MC simulations are normalized to data luminosity using cross section values in Table 2.

In the data sample, it contains over 90% of $t\bar{t}$ events. Other SM processes such W+Jets, DY+Jets, Diboson, and single top production are rather small. However, even in $t\bar{t}$ events, the leading two light jets are not always originating from W hadronic decays. In the MC simulation, we define this type of $t\bar{t}$ events, where at least one of the light jet is not from the W decay, to be "combinatorial" $t\bar{t}$ events. The way we determine the W mass peak is as follows

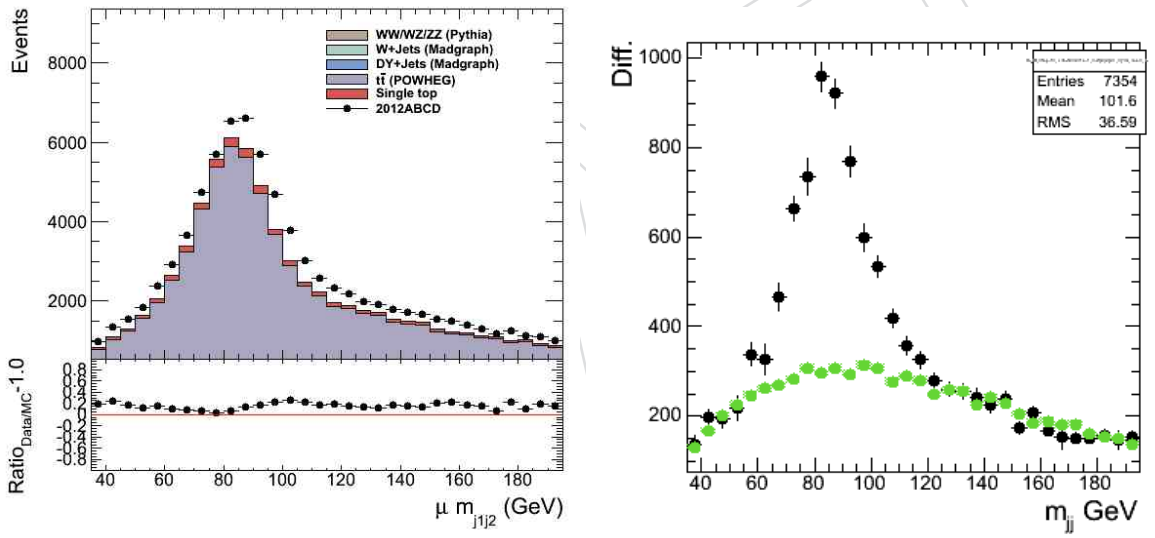


Figure 30: Left) Reconstructed W mass peak in selected "muon+≥2 b-tagged jets + ≥2 light jets" sample. The data used is the SingleMu primary dataset. MC simulations are normalized to data luminosity using production cross sections. Right) Background (DY+Jets/W+Jets/Diboson/Single top) subtracted W mass peak. The green histogram shows the estimated "combinatorial" $t\bar{t}$ background. Only the events with number of vertexes [11, 13) are included. See text for more discussions.

- First, we subtract all non-$t\bar{t}$ contributions using MC simulations from the data. Corrections due to PU weighting, lepton efficiencies, b-tagging scale factors, have been applied on MC simulations.
- We then determine the "combinatorial" $t\bar{t}$ contribution in data. We take the "combinatorial" $t\bar{t}$ contribution from MC simulation after applying all corrections and normalized using cross section. We fit the background subtracted W mass distri-

bution with a Gaussian and the "combinatorial" $t\bar{t}$ contribution. A scale factor to the "combinatorial" $t\bar{t}$ contribution can be determined ("fitted"). We also determine another scale factor for the "combinatorial" $t\bar{t}$ contribution by normalizing the MC prediction to expected yields in a sideband $125 < m_{jj} < 155\,\text{GeV}$ ("sideband"). This two scale factors have to be consistent, otherwise the scale factor from sideband normalization is used. Fig. 30 shows the background subtracted W mass distribution, where the scaled "combinatorial" $t\bar{t}$ contribution is shown.

- We further subtract the "combinatorial" $t\bar{t}$ events from data and the resulting W mass distribution is fitted with a single Gaussian to determine the W mass peak and width.
- The $t\bar{t}$ MC simulation is treated as "pseudo-data", and above procedure has been repeated to determine the W mass peak and width in MC simulation.

We divide data and MC simulation into 8 subsets in bins of number of vertexes, [0, 7, 9, 11, 13, 15, 17, 21, 35] and Fig. 31 shows the W mass peak in selected "muon+$\geq$2 b-tagged jets + $\geq$2 light jets" sample for both data and MC simulation. The data used is the SingleMu primary dataset. All other background, including "combinatorial" $t\bar{t}$ contribution, have been subtracted. Only the events with number of vertexes [11, 13) are included. The fitted W mass peak in data and MC simulation as a function of number of vertexes is also shown.

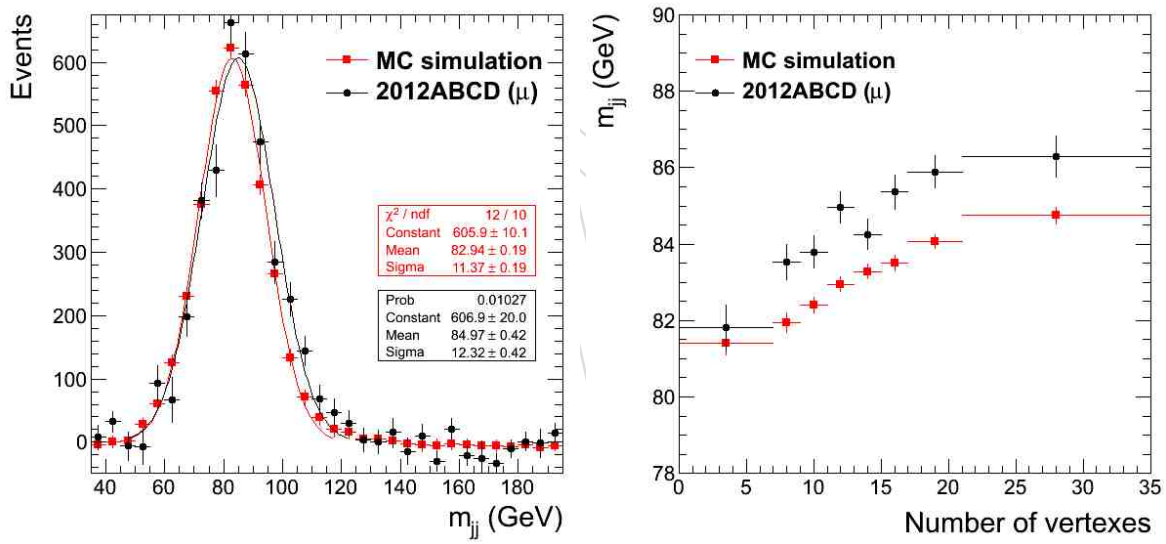


Figure 31: Left) W mass peak in selected "muon+$\geq$2 b-tagged jets + $\geq$2 light jets" sample for both data and MC simulation. The data used is the SingleMu primary dataset. All other background, including "combinatorial" $t\bar{t}$ contribution, have been subtracted. Only the events with number of vertexes [11, 13) are included. Right) The fitted W mass peak in data and MC simulation as a function of number of vertexes.

# C Yields between 2011 and 2012 datasets

work in progress

# D Other cross checks - 2012 analysis

## D.1 Evolving of $m_{lljj}$ mass spectrum in runs

Fig. 32 shows the evolving of $m_{lljj}$ mass spectrum in Run2012AB, Run2012ABC, and Run2012ABCD datasets. The cumulative distributions are shown here. All $m_{lljj}$ mass spectra show similar distributions with "excess" around 1 TeV. With more statistics, the "excess" persists.

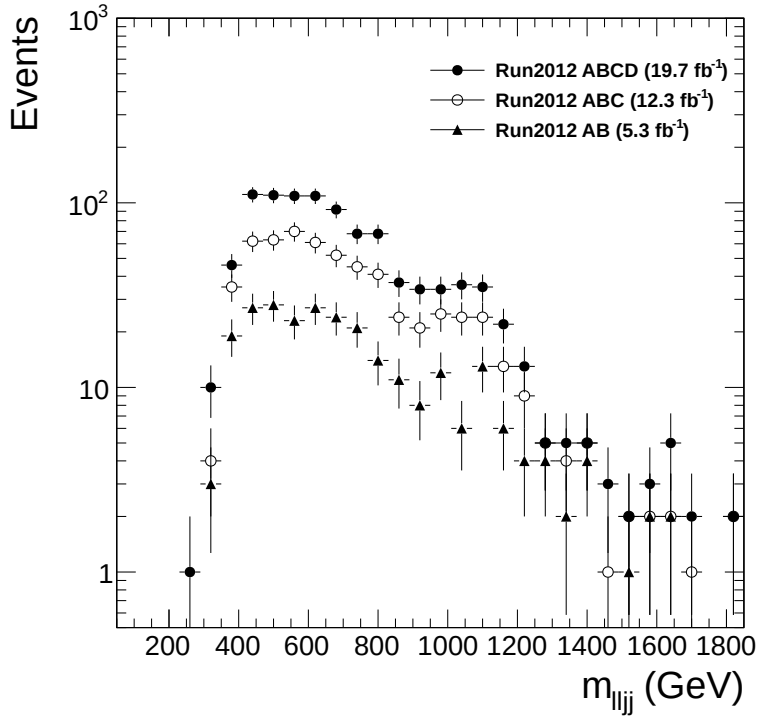


Figure 32: Evolving of $m_{lljj}$ mass spectrum in Run2012AB, Run2012ABC, and Run2012ABCD datasets.

## D.2 Relaxing the selections of the 3rd jet or b-jet veto

Fig. 33 shows the $m_{lljj}$ mass spectra in $\mu\mu$ above Z-pole channel after removing the requirements on the 3rd jet and b-jet veto, respectively. We relax these two selections one at a time and all other selections have been applied. MC simulations are normalized to data luminosity using the cross section values in Table 2. Both distributions show indications of "excess" in the mass region around 1 TeV region. However, the data/MC ratios tends to be less pronounced.

# E Background subtracted kinematic distributions - 2012 analysis

Here we show some background subtracted kinematic distributions for the "excess" events. The convention we take here is that leptons and jets are sorted in $p_{\mathrm{T}}$ and labeled as $l1$, $l2$ and $j1$, $j2$, $j3$, $j4$, respectively.

Fig. 34 shows the $p_{\mathrm{T}}$, $\eta$, and charge distributions for the leading and next-to-leading muons for the "excess" events after subtracting the known SM background contributions.

Fig. 35 shows the kinematic variables related to the dimuon system for "excess" events, after

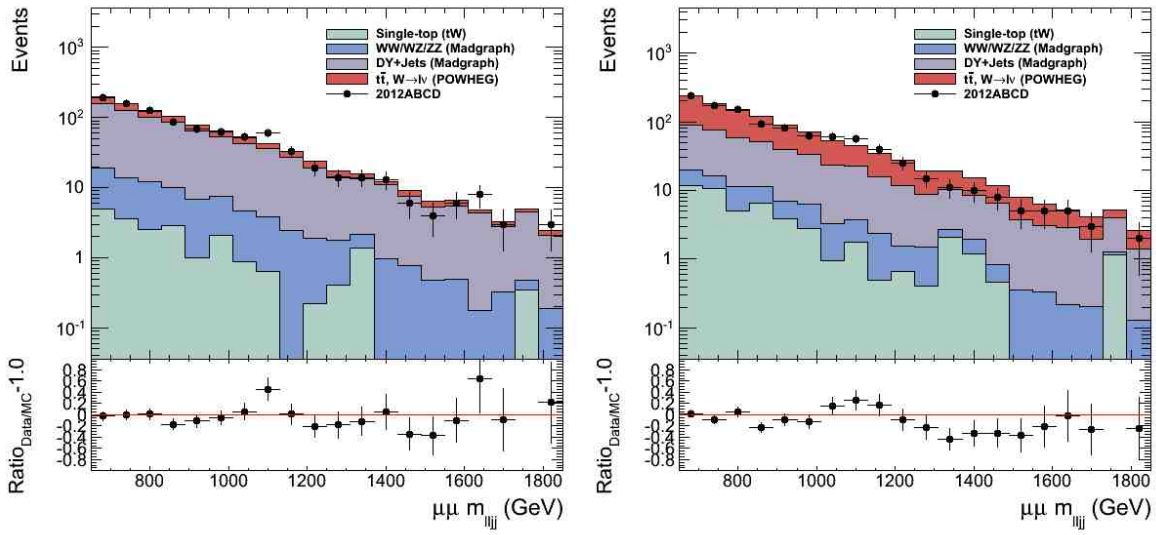


Figure 33: The $m_{lljj}$ mass spectra in $\mu\mu$ above Z-pole channel after removing the requirements, Left) on the 3rd jet and Right) b-jet veto, respectively. MC simulations are normalized to data luminosity using the cross section values in Table 2.

subtracting known SM contributions. The $\Delta\Phi$ is defined in the transverse plane of z-axis, and the polar angle $\theta_{CS}$ of $\mu^-$ is calculated in the Collins-Soper frame.

Fig. 36 shows the leading jet, next-to-leading jet, and 3rd jet $p_T$ and $\eta$ distributions for "excess" events, after subtracting known SM contributions.

Fig. 37 shows the PF-met distribution for "excess" events, after subtracting known SM contributions.

Fig. 38 shows 5-body invariant mass $m_{ll3j}$ and 6-body invariant mass $m_{ll4j}$ for "excess" events after subtracting known SM contributions.

Fig. 39 and Fig. 40 shows 3-body invariant mass and 2-body invariant mass for "excess" events after subtracting known SM contributions.

# References

[1] J. Butler et al., "Search for resonances in $ll + jj$ final states at $\sqrt{s} = 7$ TeV", CMS Analysis Note 2011/383, 2011.

[2] J. Butler et al., "Update on the Search for Resonances in $ll + jj$ Final States", CMS Analysis Note 2012/260, 2012.

[3] P. Tan et al., "Summary of the $\mu\mu jj$ anomaly", CMS Internal Note 2012/010, 2012.

[4] S. Chatrchyan et al., "Search for Heavy Neutrinos and $W_R$ Bosons with Right-Handed Couplings in a Left-Right Symmetric Model in $pp$ Collisions at $\sqrt{s}$=7 TeV", *Phys. Rev. Lett.* **109** (Dec, 2012) 261802, doi:10.1103/PhysRevLett.109.261802.

[5] S. Chatrchyan et al., "Search for pair production of first- and second-generation scalar leptoquarks in $pp$ collisions at $\sqrt{s}$=7 TeV", *Phys. Rev. D* **86** (Sep, 2012) 052013, doi:10.1103/PhysRevD.86.052013.

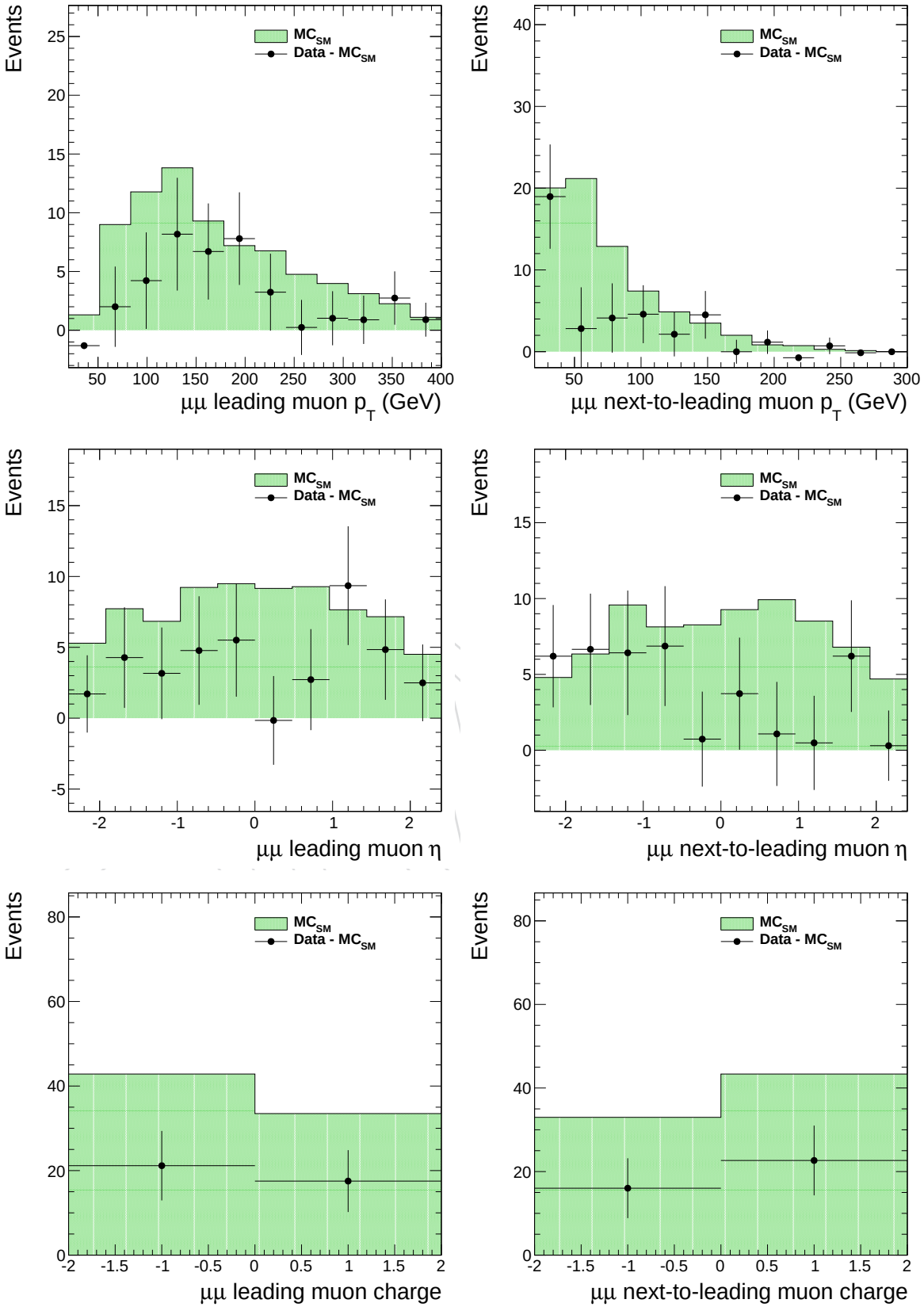


Figure 34: Left) leading muon, Right) next-to-leading muon $p_{\mathrm{T}}$, $\eta$, and charge distributions for "excess" events, after subtracting known SM contributions.

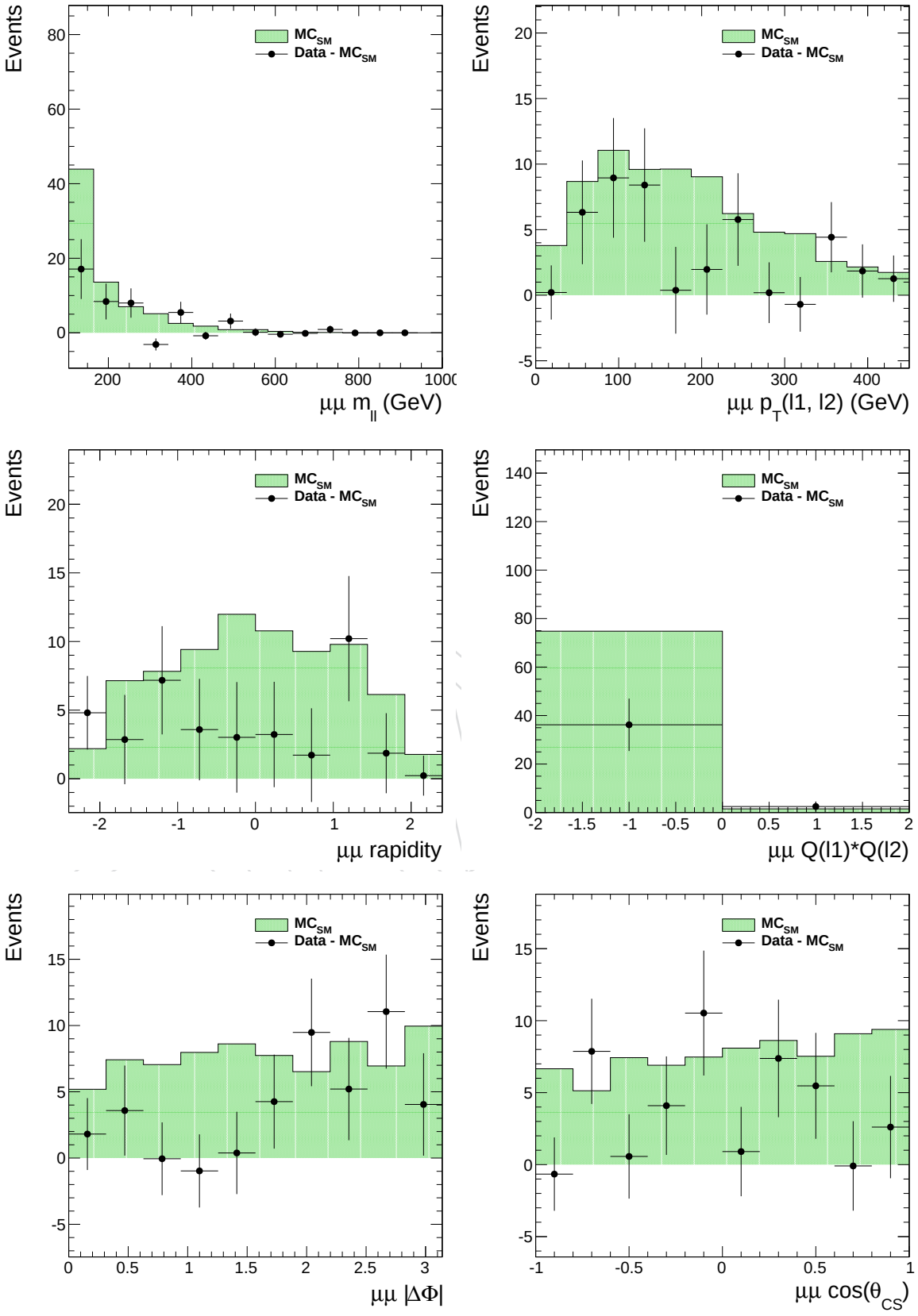


Figure 35: Kinematic variables related to the dimuon system for "excess" events, after subtracting known SM contributions. The $\Delta\Phi$ is defined in the transverse plane of z-axis, and the polar angle $\theta_{CS}$ of $\mu^-$ is calculated in the Collins-Soper frame.

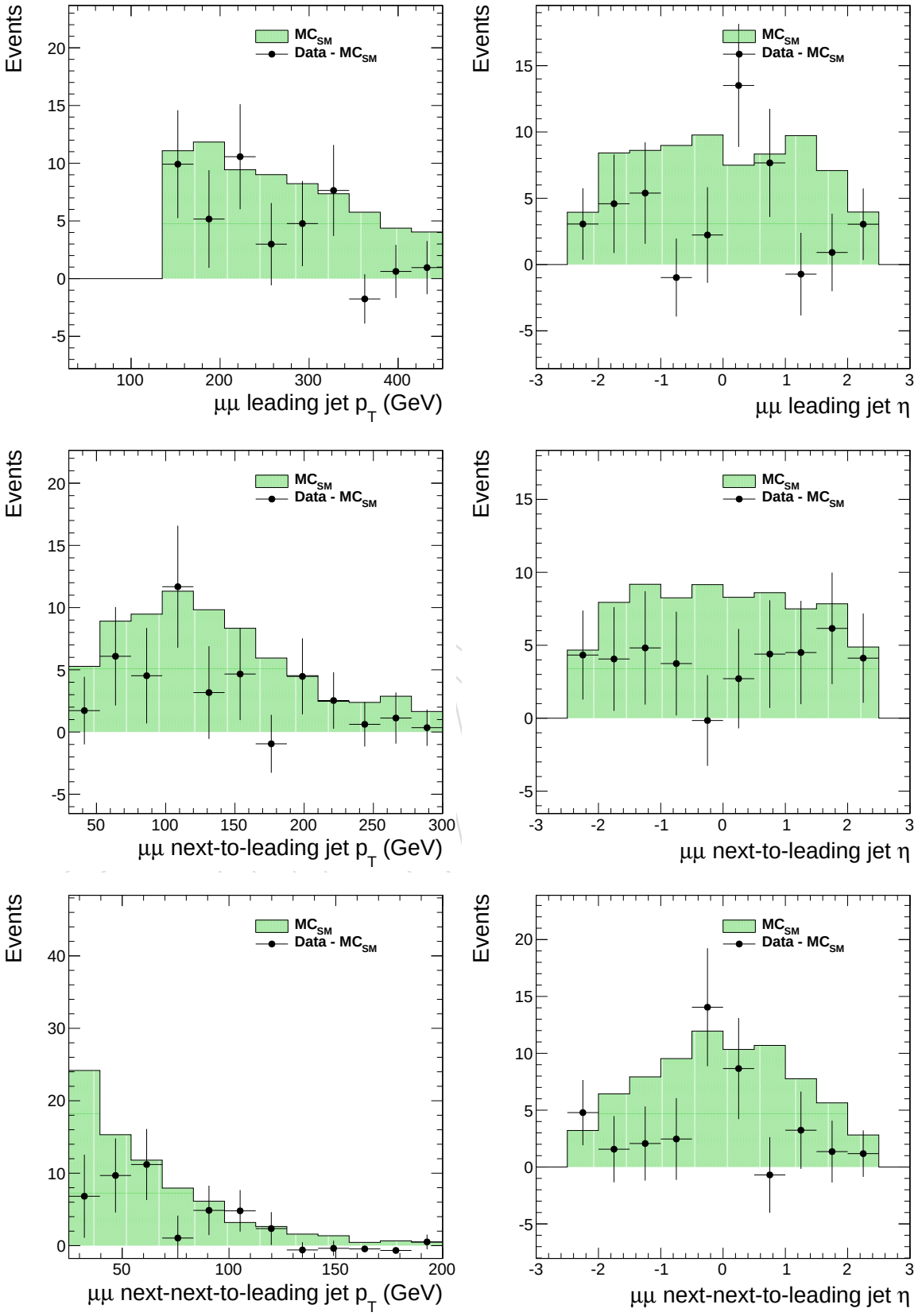


Figure 36: Top) leading jet, Middle) next-to-leading jet, Bottom) 3rd jet $p_{\mathrm{T}}$ and $\eta$ distributions for "excess" events after subtracting known SM contributions.

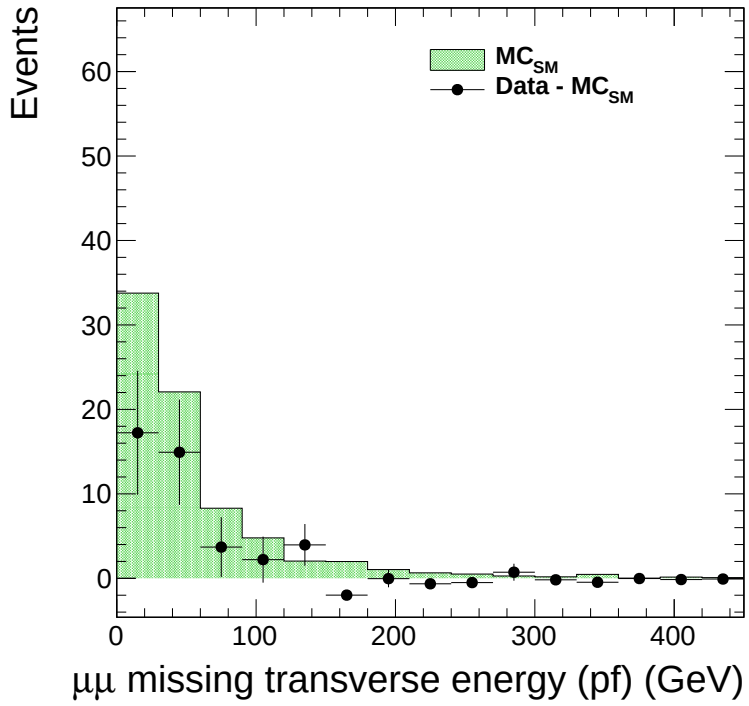


Figure 37: The PF-met distribution for "excess" events after subtracting known SM contributions.

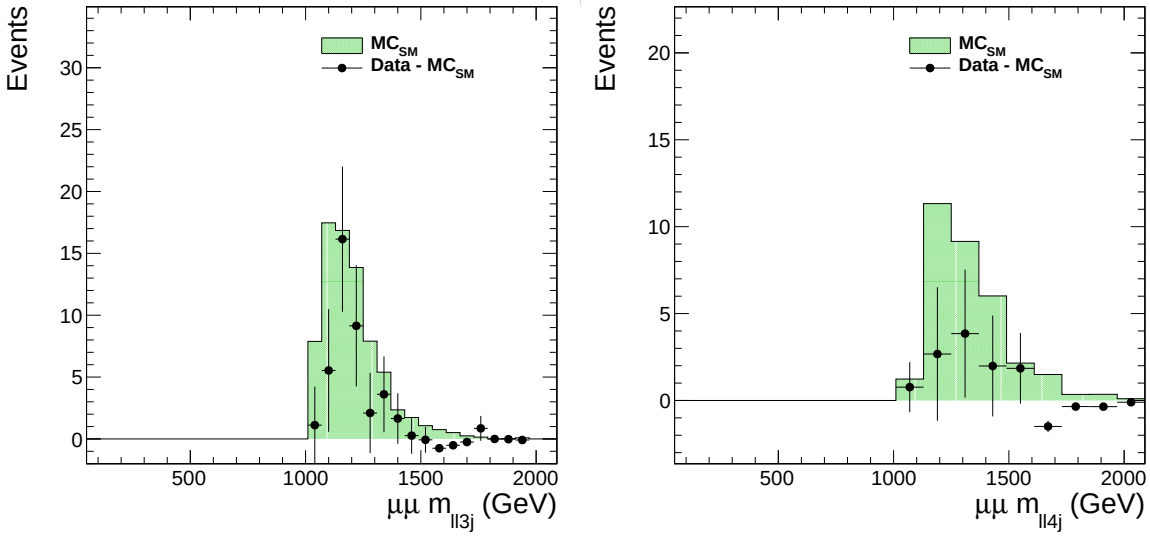


Figure 38: Left) 5-body invariant mass $m_{ll3j}$ and Right) 6-body invariant mass $m_{ll4j}$ for "excess" events after subtracting known SM contributions.

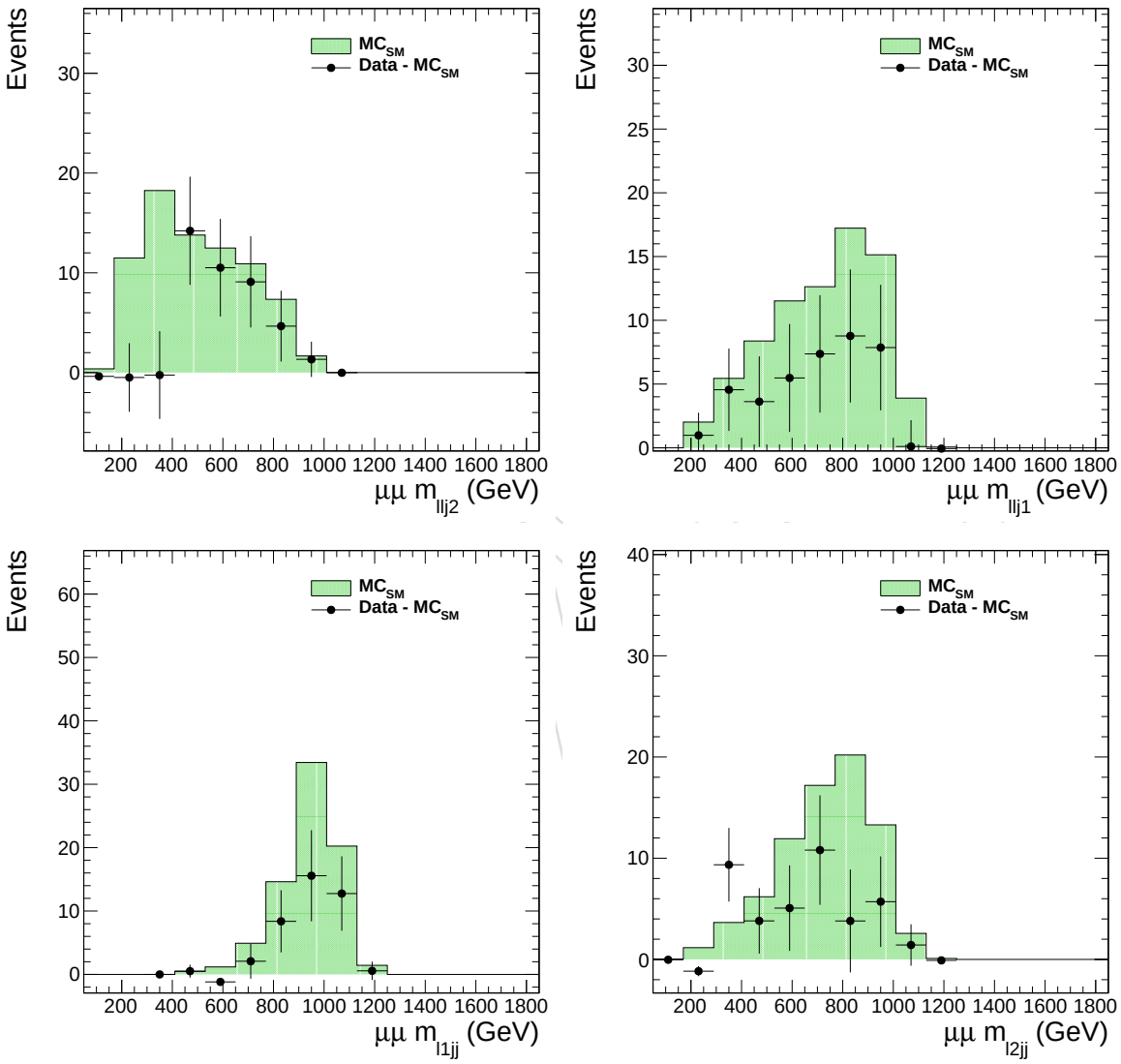


Figure 39: The 3-body invariant mass for "excess" events after subtracting known SM contributions.

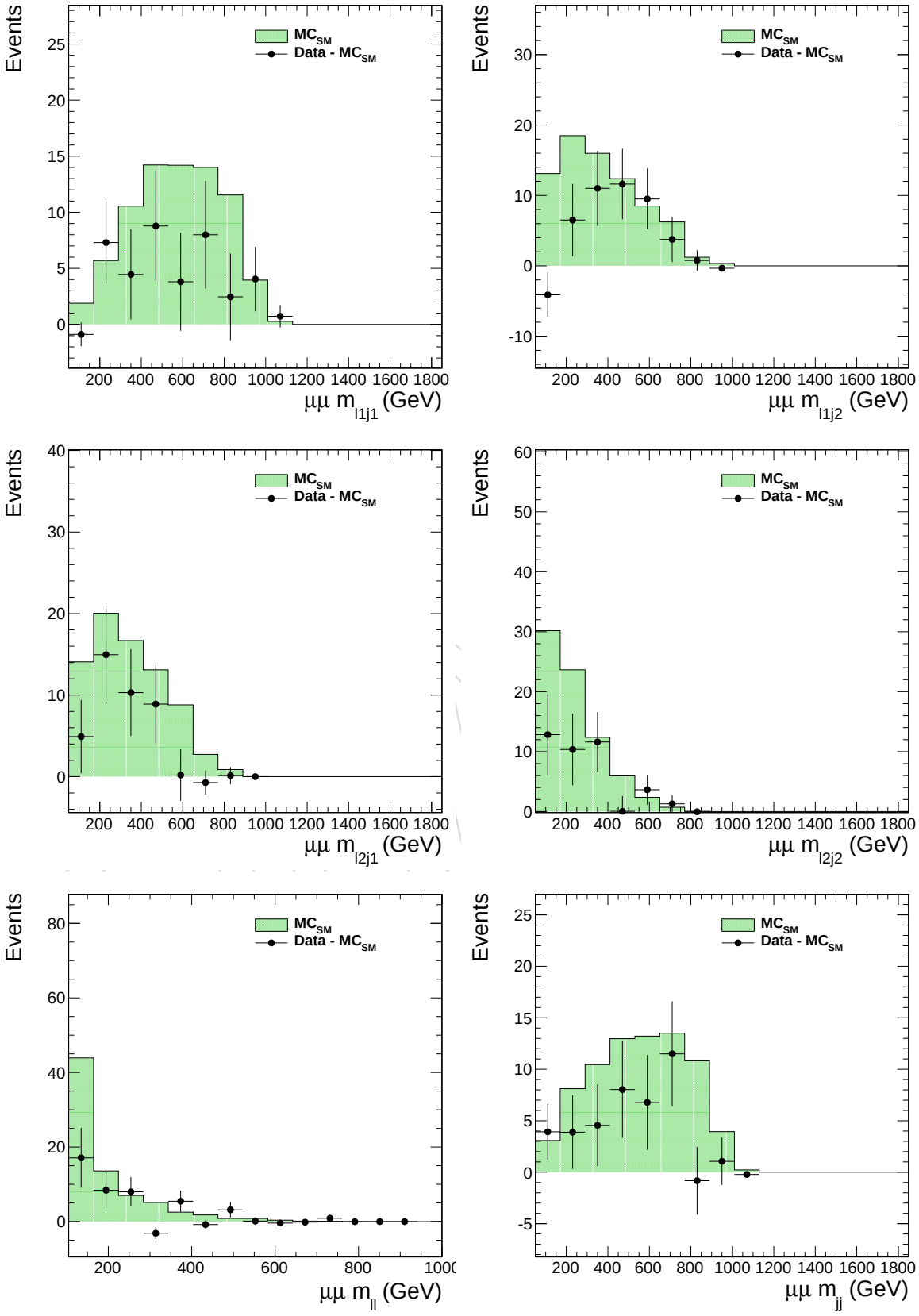


Figure 40: The 2-body invariant mass for "excess" events after subtracting known SM contributions.

[6] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[7] H.-L. Lai et al., "New parton distributions for collider physics", *Phys. Rev.* **D82** (2010) 074024, doi:10.1103/PhysRevD.82.074024, arXiv:1007.2241.

[8] J. Alwall et al., "MadGraph 5 : Going Beyond", *JHEP* **1106** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.

[9] J. Pumplin et al., "New generation of parton distributions with uncertainties from global QCD analysis", *JHEP* **07** (2002) 012, doi:10.1088/1126-6708/2002/07/012, arXiv:hep-ph/0201195.

[10] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 physics and manual", *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[11] Z. Was, "TAUOLA the library for tau lepton decay, and KKMC / KORALB / KORALZ /... status report", *Nucl.Phys.Proc.Suppl.* **98** (2001) 96–102, doi:10.1016/S0920-5632(01)01200-2, arXiv:hep-ph/0011305.

[12] Particle Data Group Collaboration, "Review of Particle Physics", *Phys. Rev. D* **86** (2012) 010001, doi:10.1103/PhysRevD.86.010001.

[13] https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation53XReReco.

[14] Y. Li and F. Petriello, "Combining QCD and electroweak corrections to dilepton production in FEWZ", arXiv:1208.5967.

[15] CMS Collaboration Collaboration, "Measurements of differential and double-differential Drell-Yan cross sections in proton-proton collisions at $\sqrt{s} = 8$ TeV", arXiv:1412.1115.

[16] CMS Collaboration, "Measurement of Z production as a function of $p_{\mathrm{T}}$, rapidity", *CMS PAS* **SMP-13-013** (2014).

[17] CMS Collaboration, "Measurement of the differential top-quark pair production cross section in the dilepton channel at 8 TeV", *CMS PAS* **TOP-12-028** (2014).

[18] https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideHiggsAnalysisCombinedLimit?skin=drupalHow_to_prepare_the_datacard.